The impact of single nucleotide polymorphisms in human genes that regulate hepcidin and iron on oral iron absorption and the risk of anaemia in Africans

Brief title: Genes-in-Action iron study

Momodou W. Jallow

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy (PhD) Of the University of London

February 2021

Department of Infection Biology
Faculty of Infectious and Tropical Disease
LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Declaration

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

DECLARATION OF OWN WORK

All students are required to complete the following declaration when submitting their thesis.

Please note: Assessment misconduct includes any activity that compromises the integrity of your research or assessment of is will be considered under the Assessment Irregularity Policy. This includes plagiarism, cheating and failure to follow correct progression and examination procedures.

Please see the following documents for further guidance:

- Research Degrees Handbook
- Assessment Irregularities Policy

Supervisors should be consulted if there are any doubts about what is permissible.

1. STUDENT DETAILS

Student ID Number	151342	**Title**	Mr
First Name(s)	Momodou W.		
Surname/Family Name	Jallow		
Programme of Study	Doctor of Philosophy		
LSHTM Email (if this is no longer active, please provide an alternative)	momodou.jallow@lshtm.ac.uk		

2. TITLE OF THESIS

Title of Thesis	The impact of single nucleotide polymorphisms in human genes that regulate hepcidin and iron on oral iron absorption and the risk of anaemia in Africans

3. DECLARATION

I have read and understood the LSHTM's definition of plagiarism and cheating. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

I have read and understood the LSHTM's definition and policy on the use of third parties (either paid or unpaid) who have contributed to the preparation of this thesis by providing copy editing and, or, proof reading services. I declare that no changes to the intellectual content or substance of this thesis were made as a result of this advice, and, that I have fully acknowledged all such contributions.

I have exercised reasonable care to ensure that the work is original and does not to the best of my knowledge break any UK law or infringe any third party's copyright or other intellectual property right.

Student Signature	
Date	13 October 2020

Dedication

I dedicate this work to my beloved Mum, Ousainatou Jallow who made sure that I go to school even if it meant selling her jewellery and clothes to pay my school fees, and my late Dad Mamadou Kally Jallow, who did not live to witness the fruits of his hard work in making sure that I went to school.

Acknowledgement

My PhD journey has been an incredibly exciting adventure with four years filled with challenges, hope, experience and interactions with several inspiring people. Some of these people have moved on to other places while others are still within the MRC or LSHTM.

First, I express my sincere gratitude to Professor Andrew Prentice for believing in me and providing me the financial support to pursue this PhD. You have shown strong confidence in me and I am highly grateful. Also, my special thanks go to Dr Carla Cerami, my Associate supervisor, who is always there to make sure I have the necessary supervisory support on the ground at MRCG at LSHTM. Thank you for all the support for making this PhD a success. I also, extend my sincere gratitude to Dr Susana Campino, my Supervisor at the LSHTM for always being there to answer my questions, and for the helpful guide throughout this PhD journey.

My thanks also go to my initial supervisor, Dr Branwen Hennig who introduced me to the human genetic world. You have created a lasting impact on my career despite the short duration we had together. You are an exceptional person who has laid a solid foundation for me to continue through the PhD journey. Your made sure that I am in great hands before you left.

My sincere gratitude to Dr Chrissy Roberts, who introduced me to R statistical software. You strengthened my molecular genetics analytical skills which had immensely impacted my laboratory skills.

I am also, extending my sincere thanks to Jenny Howard and Joanne Tapper both were former (LSHTM/ING) Administrators. These two ladies made the beginning of my PhD journey smooth by providing all the administrative support that made my first LSHTM visits both exciting and fun. You left a lasting positive impression in me.

Through your support, I was able to settle down very well into the PhD. I am also thankful to the administrative staff at the MRCG at LSHTM Learning and Development Department (both past and present): Mrs Mariatou Sallah, Mr Ismaila Danso and Ms Adam Drammeh. These people immensely smoothen my PhD journey by giving me all the administrative support I needed.

Furthermore, I would like to thank Dr Amat Bah, for the moral and intellectual advices. You gave me very important guidance from the beginning of this PhD to the end. Also, I am thankful to Dr James Cross, Fatou Joof, Dr Philip James and Dr Toby Candler for the peer support during my stay in Keneba. Interacting with you lightened the PhD stress. The PhD pancakes were great!

To the Keneba Field station Administrative staff, Fatou Colley, Kanimang Touray and Buba Jabang, I say a big thanks to you all! You made my Keneba stay smooth.

My special, thanks go to the Genes-in-Action Iron study team: Nuha Camara, Dembo Danso, Lamin Jatta, Buba Sonko, Edrissa Sinjanka, Babatam Bah, Morris Ndene, Lamin Darboe, Bakary Sonko, Kabiru Sise and Alasana Saidykhan. Without your support, the GiA study wouldn't have happen. Your hard work and dedication made it easy for me to conduct this study with great success.

I am also thankful to Dr Modou Jobe, Bakary Darboe Isatou Dibba, my Platt building Family at Fajara. I strongly appreciate the prayers and moral support. The social events and discussions on current events lightened my burden. You made my Fajara stay smooth and enjoyable.

Finally, extent special thanks to my family for the patients on the numerous travels and and long nights awake. Your prayers are valuable.

Above all, I thank the Almight Allah (God) for giving me the health and strength to be able to successfully complete this Journey.

Abstract

Background: Up to 60% of women and children living in low- and middle-income countries (LMICs) are anaemic. Food fortification and iron supplementation are the most common measures employed to combat anaemia. However, these are not effective treatments for anaemias caused by non-nutritional factors. Genome-wide association studies (GWAS) mainly in Europeans and Asians have identified single nucleotide polymorphisms (SNPs) within the hepcidin and iron regulatory genes that are associated with the risk of anaemia. Several of these SNPs are in the *TMPRSS6* gene, which encodes matriptase-2, a protein that regulates the expression of hepcidin. This thesis examined the impact of SNPs in the iron regulatory genes previously reported in non-African populations, on the risk of anaemia and on impaired oral iron absorption in Africans.

Methods: First, the literature was searched for genetic variants identified in the hepcidin and iron regulatory genes, that are associated with low iron status. Second, we investigated the effects of common *TMPRSS6* and transferrin (*TF*) SNPs on iron status in a cohort of healthy individuals from rural Gambia (n=1315). Third, a recall-by-genotype (RbG) study was conducted to investigate the impact of carrying single or multiple alleles at the common *TMPRSS6* SNPs on oral iron absorption in healthy individuals from rural Gambia.

Results: *TMPRSS6* rs855791, rs4820268 and rs2235321, and *TF* rs3811647 are the most common SNPs that associated with low iron status. We did not find effects of any of the *TMPRSS6* SNPs on the risk of anaemia. However, we found that *TMPRSS6* rs2235321 was associated with serum hepcidin concentration, with a more substantial effect on individuals with low haemoglobin or ferritin. Also, *TF* rs3811647 had a significant influence on transferrin and its binding capacity,

with a single allele effect of 8-12%. In the RbG study, we did not find any effect of the three *TMPRSS6* SNPs on oral iron absorption. However, we found that each of the *TMPRSS6* SNPs affects hepcidin, with carriers of major alleles having higher hepcidin compared to minor allele carriers. Also, we found that heterozygotes at both rs2235321 and rs855791 did not alter their hepcidin concentration after an oral iron dose, whereas, individuals in all the other genotype groups did.

Conclusions: This thesis confirms the previously observed association between the *TF* rs3811647 and transferrin in other Africans and Europeans replicates in West Africans. However, we could not demonstrate that the previous associations between *TMPRSS6* gene variants and iron status, exist in West Africans. This lack of replication might be due to the high genetic diversity that exists in African populations. We identified an effect of *TMPRSS6* rs2235321 on serum hepcidin concentration. In the RbG study, the three *TMPRSS6* SNPs studied influenced serum hepcidin levels but not oral iron absorption in healthy individuals. This finding suggests that there might be an alternate pathway of iron regulation independent of hepcidin at the enterocytes. These findings highlight the need to conduct more research on genetic determinants of iron status in African populations. Investigating more genetic markers and in different populations may provide a clearer insight into the role of genetic risk factors on iron deficiency and anaemia in African populations. Identifying the role of genetic risk factors of iron status may pave the way for the formulation of population-specific anaemia control measures.

Key terms: Anaemia; Iron deficiency; SNPs; TMPRSS6; TF; Hepcidin, Hepcidin regulatory genes; African populations; iron absorption.

Table of Contents

Table of Figures

Abbreviations

AGP	Alpha-1 glycoprotein
AI	Anaemia of inflammation
ANOVA	Analysis of variance
ARS	Allele risk score
BMI	Body mass index
BMP	Bone morphogenetic protein
BMPR	Bone morphogenetic protein receptor
CRP	C-reactive protein
DfID	Department for International Development
DMT1	Divalent metal transporter 1
DNA	Deoxyribonucleic acid
EDTA	Ethelenediametelenetatraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ERK	Extracellular signal-regulated kinase
eQTL	Expression quantitative trait loci
G6PD	Glucose-6-phosphatase dehydrogenase
GAM	The Gambia
GiA	Genes-in-Action
GWAS	Genome-wide association study
HAMP	Hepatic antimicrobial peptide
Hb	Haemoglobin
HBA	Haemoglobin subunit alpha
HBB	haemoglobin subunit beta
HbS	Haemoglobin S
Hct	Haematocrit
HCP1	Haem carrier protein 1
HFE	Human haemochromatosis protein
HH	Hereditary haemochromatosis
HJV	Haemojuvelin
HUGE	Human genetic epidemiology
H3Africa	Human Hereditary & Health in Africa

ID	Iron deficiency
IDA	Iron deficiency anaemia
IRIDA	Iron-refractory iron deficiency anaemia
IL6	Interleukin 6
IL6R	Interleukin 6 receptor
JAK	Janus associated kinase
KWDSS	Kiang West Demographic Surveillance System
LMIC	Low- and middle-income Countries
LSHTM	London School of Hygiene & Tropical Medicine
MAF	Minor allele frequency
MAPK	Mitogen-associated activated protein kinase
MCH	Mean corpuscular haemoglobin
MCHC	Mean corpuscular haemoglobin concentration
MCV	Mean corpuscular volume
MNS	National Micronutrient Survey
MRCG	Medical Research Council The Gambia Unit
NHANES	National Health and Nutrition Education Survey
PCR	Polymerase chain reaction
pSMAD	phosphorylated SMAD
RBC	Red blood cell
RbG	Recall-by-genotype
RDW	Red cell distribution width
RNA	Ribonucleic acid
SCC	Scientific coordination committee
SE	Standard error
sHJV	Soluble haemojuvelin
SLC11A2	Solute carrier family 11 member 2
SLC40A1	Solute carrier family 40 member 1
SMAD	Sons of mothers against decapentaplegic
SNP	Single nucleotide polymorphism
SSA	Sub-Saharan Africa
STAT	Signal transducer and activator transcription

TF	Transferrin
TFR1	Transferrin receptor 1
TFR2	Transferrin receptor 2
TIBC	Total iron binding capacity
TMPRSS6	Transmembrane protease serine 6
TSAT	Transferrin saturation
UIBC	Unsaturated iron binding capacity
WHO	World Health Organisation
WABR	West Africa BioResource
YLD	Yearly life lost in disability

Chapter 1:

Candidate's contribution and thesis structure

1.1. Candidate's contribution

Prior to starting my PhD, I have been working with the MRC International Nutrition Group, which later became the Nutrition Theme at the MRC Unit The Gambia, for eight years. I started as a laboratory technician and I rose to a Scientific Officer responsible for performing hepcidin assay optimisation and validation. Over this period, I have been involved in several research projects investigating the influences of iron and hepcidin on various diseases. As a result, I developed strong interest in iron research. This is why when this PhD opportunity was presented to me by Prof Andrew Prentice and Dr Branwen Hennig (my initial supervisor) I saw it as a perfect opportunity for developing my career in this cross-cutting field involving nutrition and genetics.

When I started the PhD, I did an extensive literature review on the topic and I made significant changes to the initial study design. These changes include introducing additional candidate genetic variants, optimising the study design and the idea of investigating the cumulative effect of SNPs on iron status.

With the departure of my initial supervisor and coming of the new supervisory team (Dr Susana Campino and Dr Carla Cerami) at the early stage of my PhD, I had to learn with them to ensure that they were familiar with my PhD topic and plans.

I spearheaded the selection of the candidate gene variants for the main study, wrote the study design and protocol, and presented the amended protocol at the MRCG Scientific Coordinating Committee (SCC), submitted the ethics application to the London School of Hygiene & Tropical Medicine via the online ethics portal (LEO online) for ethics approval. I responded to the ethics questions with the support of my supervisors, and resubmitted amended versions. Furthermore, I wrote the study information sheets and all the standard operating procedure.

I mobilised the study team (one nurse, 2 fieldworkers, one laboratory technician and one data manager) in consultation with the respective section heads. Also, I managed the purchasing of the study consumables and managed all the study supplies. I planned and conducted the pilot study (September 2016 to January 2017). I worked with the data manager to develop the study database. I selected the candidate SNPs and identified the potential participants from the genotyped cohort. I coordinated all the participant recruitment, follow-ups, study visits and all the study procedures.

I supervised sample collection, processing and storage. Also, I performed the hepcidin analysis using an enzyme-linked immunosorbent assay (ELISA). I supervised iron biomarkers' analysis which was done by a member of the Keneba Laboratory platform Team. I supervised data recording in the study forms and verified all forms before handing over to the data management team for entry into the study database.

I curated all the study data and performed all the analysis and presented the results to my supervisory team for review. I wrote all the manuscripts, managed co-authors' comments and reviews, and incorporated them in the final manuscripts, and managed journal submissions. I ensured that all the authors approved the final manuscripts before submission for publication.

1.2. Thesis scope and composition

The purpose of this thesis was to investigate the genetic determinants of anaemia and their effects on the response to oral iron supplementation in Africans. Anaemia is a global health problem, and women and children living in low- and middle-come countries are the most vulnerable [1]. High burden of anaemia remains in these settings despite decades of implementing anaemia control measures. Current control measures have been mainly targeted at alleviating anaemia caused by nutritional iron

deficiency [2,3]. However, only about half of anaemia cases are attributed to nutritional iron deficiency, and this varies across populations and geographical regions [4,5]. The aetiology of anaemia is complex, and the role of genetic risk factors has not been thoroughly investigated, particularly in West African populations, where the burden of anaemia is among the highest.

The discovery of hepcidin brought in a new understanding of iron biology and metabolism [6]. Subsequently, genome-wide association studies, mainly on European and Asian populations, identified several genetic variants that are associated with impaired iron status and the risk of anaemia [7–10]. The studies constituting this thesis were designed to investigate whether genetic variants in the hepcidin and iron regulatory genes predispose healthy Africans to anaemia and whether such variants would impede the response to oral iron supplementation. This work was made possible by the existence of the Keneba Biobank at the MRCG at LSHTM, which comprises of individuals with genotype and phenotype data. This resource enabled the recall of participants for subsequent detailed investigations based on their previously available genotype data. This thesis consists of chapters with published papers incorporated into chapters and papers not published are yet presented in the format they were submitted for publication.

The thesis structure is as follows:

Chapter 1: Candidate's contribution and thesis structure

Chapter 2: This chapter explains the background and rationale for conducting the PhD.

Chapter 3: This chapter is a review paper on the genetic determinants of iron imbalance among global populations. It also presents results of analysis of genetic data from the 1000genomes project and the Keneba Biobank population. It describes differences in allele frequencies and linkage disequilibrium and investigates signatures of selection across global populations. This was done to enable the selection of candidate gene SNPs for the subsequent studies. This is presented in a paper titled: *Differences in the frequency of genetic variants associated with iron imbalance among global populations,* that has been published in PLOS ONE (https://dx.plos.org/10.1371/journal.pone.0235141).

Chapter 4: This chapter is a research paper investigating the effects of common *TMPRSS6* and *TF* genes' SNPs on iron status in healthy individuals of all ages, titled: *Association of common TMPRSS6 and TF gene SNPs with hepcidin and iron status in healthy rural Gambians***.** This paper describes the effects of carrying single or multiple risk alleles at the common *TMPRSS6* and *TF* SNPs on the risk of anaemia. This paper is has been submitted to Nature Scientific Reports.

Chapter 5: This chapter is the published protocol paper. This paper described the background of the recall-by-genotype study. In this paper we presented the study design, study location, ethical permissions, informed consent and confidentiality, sample size estimation and the statistical analyses methods would be applied. The paper is published in F1000 Research (https://f1000research.com/articles/8-701/v1).

Chapter 6: This chapter describes the results of the recall-by-genotype study. This results have been written in a paper titled: *Common variants in the transmembrane protease serine 6 (TMPRSS6) gene alter hepcidin but not oral iron absorption in*

healthy Gambian adults: a recall-by-genotype study, and submitted to Current Development in Nutrition.

Chapter 7: This chapter presents the summary discussion and conclusions including limitations of the thesis. Furthermore, the public health implications and future research needs are discussed.

Chapter 8: Appendices: This chapter consists of documentation of the study and photos taken during the PhD.

1.3. Supervisory team

My supervisors for this PhD are Dr Susana Campino, ITD Faculty, LSHTM and Dr Carla Cerami, Nutrition Theme, MRCG at LSHTM.

The Advisory committee members are: Prof Andrew Prentice and Dr Robert Butcher.

1.4. Funding

This PhD is funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID), under the MRC/DFID Concordat agreement to the MRC International Nutrition Group (MRC-ING), grant MC-A760-5QX00.

1.5. References

1. Kassebaum, N. J. The Global Burden of Anemia. *Hematol. Oncol. Clin. North Am.* **30**, 247–308 (2016).

2. World Health Organization. *Essential nutrition actions: improving maternal, newborn, infant and young child health and nutrition.* (World Health Organization, 2013).
3. Howson, C. P., Kennedy, E. T. & Horwitz, A. *Prevention of Micronutrient Deficiencies: Tools for Policymakers and Public Health Workers.* (National Academies Press, 1998).
4. Le, C. H. H. The Prevalence of Anemia and Moderate-Severe Anemia in the US Population (NHANES 2003-2012). *PLOS ONE* **11**, e0166635 (2016).
5. Petry, N. *et al.* Micronutrient Deficiencies, Nutritional Status and the Determinants of Anemia in Children 0–59 Months of Age and Non-Pregnant Women of Reproductive Age in The Gambia. *Nutrients* **11**, 2275 (2019).
6. Camaschella, C., Nai, A. & Silvestri, L. Iron metabolism and iron disorders revisited in the hepcidin era. *Haematologica* **105**, 260–272 (2020).
7. Li, J. *et al.* GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children. *Hum. Mol. Genet.* **22**, 1457–1464 (2013).
8. Kullo, I. J., Ding, K., Jouni, H., Smith, C. Y. & Chute, C. G. A Genome-Wide Association Study of Red Blood Cell Traits Using the Electronic Medical Record. *PLoS ONE* **5**, e13011 (2010).
9. McLaren, C. E. *et al.* Genome-Wide Association Study Identifies Genetic Loci Associated with Iron Deficiency. *PLoS ONE* **6**, e17390 (2011).
10. Gichohi-Wainaina, W. N. *et al.* Inter-ethnic differences in genetic variants within the transmembrane protease, serine 6 (TMPRSS6) gene associated with iron status indicators: a systematic review with meta-analyses. *Genes Nutr.* **10**, 442 (2015).

Chapter 2:
General Introduction

Chapter description

This chapter discussed the rational for conducting, the literature review, and aims and objectives the PhD thesis.

2.1. Background and Rationale

Iron is an essential micronutrient required for most physiological processes in both humans [1] and microorganisms [2]. In humans, the primary requirement for iron is erythropoiesis, which consumes 60-70% of body iron [3]. Iron is also required for DNA synthesis, electron transport, cell proliferation and the maintenance of effective immune system [2,4]. Thus, reduced or inadequate iron availability in the body results in impaired physiological function [5].

Microorganisms also need iron for growth and proliferation [6]. Therefore, excess body iron favours pathogens as they utilise iron to grow and proliferate [2]. Also, excess iron in the body results in the generation of reactive oxygen species which can cause organ damage [7]. In contrast, a deficit in iron supply results in iron deficiency and anaemia, which is also detrimental to health and wellbeing.

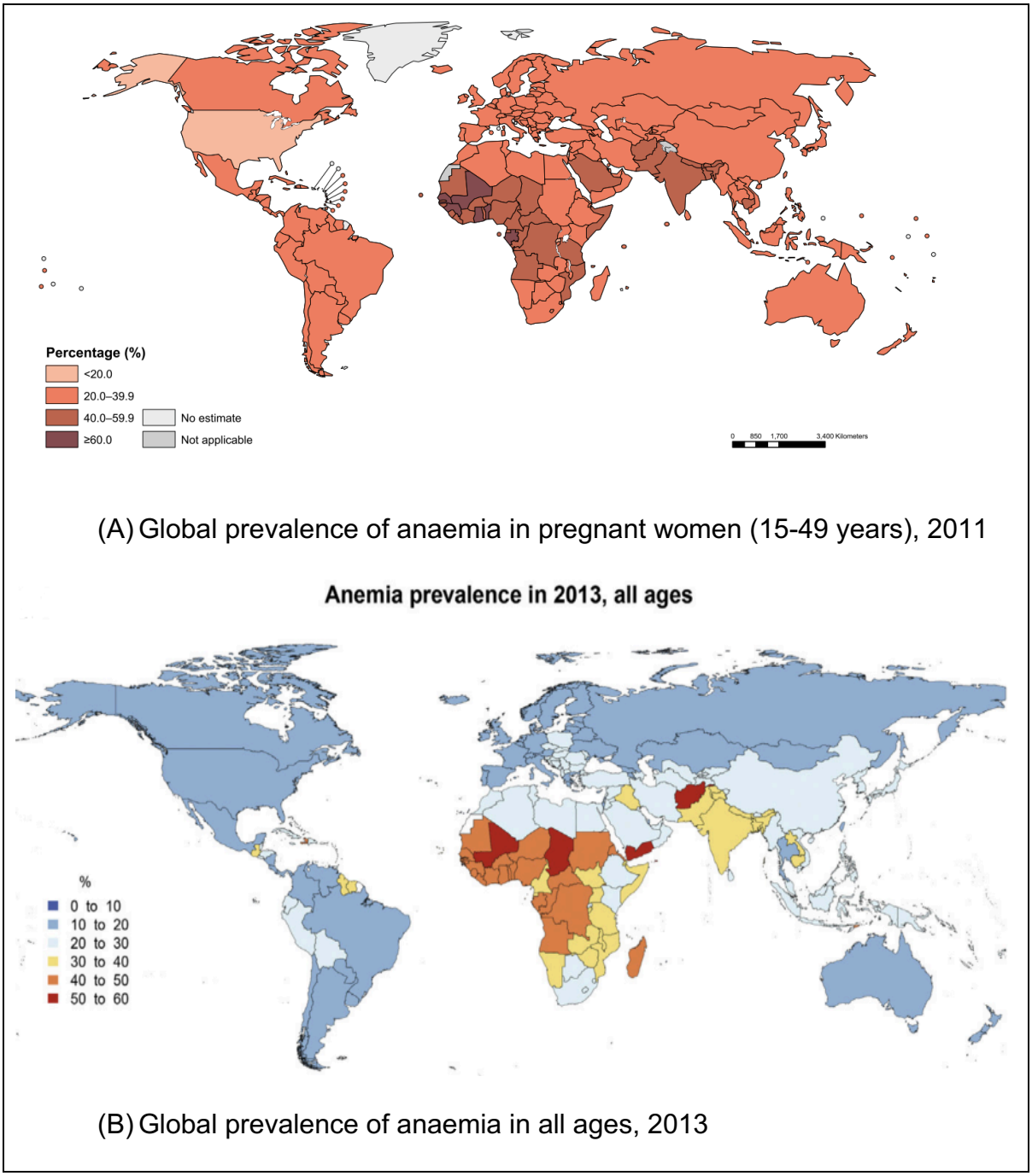
In order to avoid the consequences of either excess or deficit, the human body evolved to maintain an optimum balance [7]. However, despite effective iron regulatory mechanisms, inevitably, iron imbalance occurs. Among the two iron imbalances, anaemia (deficit in the supply of iron) is the most common nutrient disorder worldwide [8].

2.2. Epidemiology and impact of anaemia

Anaemia remains a global health concern that is responsible for approximately 8% of all non-fatal health loss from disease [9]. In 2013, anaemia was estimated to affect 1.9 billion people worldwide [9,10]. Low- and middle-income countries (LMICs) carry 89% of the global anaemia burden [9]. In sub-Saharan Africa, up to 60% of women of reproductive age are anaemic (**Figure 1A**), and the same is true for children under five years of age in this setting [9]. In Western and Central Africa, the prevalence of anaemia in the general population reached 50% (**Figure 1B**). In the Gambia, a recent

national nutritional status survey identified that 50.4% and 59.0% of pre-school children have anaemia and iron deficiency, respectively [11]. Also, in non-pregnant woman of reproductive age, 50.9% and 41.4% are anaemic and iron deficient, respectively [11]. This indicates a high burden of anaemia in the Gambia, particularly in population groups that are most vulnerable to the effects of anaemia.

Despite being amenable to cheap and widely available medicinal products (iron preparations), the overall disease burden of anaemia is higher than cardiovascular diseases, asthma and diabetes combined, resulting in 61.5 million yearly lost in disability (YLD) [9]. Also, the Global Burden of Disease Study 2016 estimated that iron deficiency anaemia (IDA) is the primary cause of YLD in women and it accounts for one-in-five leading causes of the YLD burden [9]. Anaemia has devastating consequences, particularly in young children, as it results in impaired cognitive function, which in turn impacts educational attainment [12,13]. During pregnancy, anaemia can lead to adverse outcomes in both the mothers and newborns. Also, anaemia can lead to impaired physical activity in adults, which affects working capacity and economic productivity [14]. In the elderly, anaemia leads to reduced cognition [13,15] and poor outcomes of chronic diseases [16].



(A) Global prevalence of anaemia in pregnant women (15-49 years), 2011

(B) Global prevalence of anaemia in all ages, 2013

Figure 1. Global estimates of anaemia prevalence.

(A) in pregnant women, 15 – 49 years in 2011 [17] and (B) in all ages, 2013 [18]

2.3. Aetiology of anaemia

The aetiology of anaemia is complex and multifactoral. One of the significant causes of anaemia is nutritional iron deficiency, which has been commonly attributed to approximately 50% of cases [9,19]. However, this varies across populations. For example, Petry and colleagues reported that in settings where the prevalence of anaemia is higher than 40%, only 25% of cases are attributable to iron deficiency [20]. This suggests that about 75% of anaemias in such settings may be due to reasons that are not purely nutritional iron deficiency. This information is essential in the fight against anaemia as most of the anaemia control programmes are tailored towards alleviating nutritional deficiency anaemia [21]. Other important causes of anemia are micronutrient deficiencies, particularly iron, vitamin A, folate, and vitamin B12 deficiencies. Given the different factors that influence anaemia, adjusting for these factors in different settings and populations are crucial to determine the true prevalence of anaemia. For example, in high inflammation burden settings, adjusting for biomarkers of inflammation would be useful when assessing anaemia.

Iron deficiency anaemia (IDA) usually results from a single or a combination of three main factors: (a) inadequate daily iron intake and increased demand, particularly at certain stages of life, including pregnancy and childhood; (b) chronic blood loss (e.g. menstruation, ulcers or parasitic infections); and (c) impaired iron absorption and utilization (e.g. which could be caused by genetic factors) [9,22].

Inadequate iron intake is one of the most common causes of IDA in SSA, particularly in young children and women of reproductive age [23,24]. This is due to the consumption of diets that are low in iron content or low bioavailable haem iron [24]. Meat is a major source of dietary iron; however, meat consumption is generally low in many African societies due to high cost [25]. Also, the consumption of foods that inhibit iron absorption

such as tea reduces the iron absorption [25]. Although consumption of iron absorption enhancers such as vitamin C promotes effective absorption, most African plant-based foods such as cereals and legumes contain phytates that inhibit iron absorption [25,26]. Also, most beverages contain iron-binding phenolics that limit iron absorption [25]. Therefore, in settings where the availability of iron-rich diet is low, increase demand for iron in children and women of reproductive age may increase to prevalence of anaemia in these groups.

In children, increase demand for iron occur due to rapid growth spurt, and thus, inadequate supply of dietary iron to compensate for this need my result in iron deficiency [27]. Similarly, women of reproductive age have an increasing demand for iron due to regular blood loss and childbirth [14]. In settings where dietary in iron intake is inadequate, women of reproductive may have a higher predisposition to iron deficiency anaemia [28,29]. Also, chronic blood loss can occur due to ulcers and parasitic infections. Children are most vulnerable to parasitic infection such as soil-transmitted helminths [28]. Furthermore, acute and chronic infections/inflammations such as malaria [29], tuberculosis, HIV and cancers may increase the risk of anaemia and iron deficiency anaemia [9]. Furthermore, respiratory infections in children impair iron absorption due to increase in hepcidin levels [30].

Furthermore, genetic risk factors such as haemoglobinopathies (sickle cell and thalassaemias) are known to reduce Hb levels [31]. Also, genetic defects in the DMT1 or the transmembrane protease serine 6 (TMPRSS6) (the gene that encodes matriptase 2) genes are associated with impaired iron absorption [32,33]. DMT1 plays a crucial role in iron absorption as it transfers dietary iron into the duodenal enterocytes [34]. Similarly, the *TMPRSS6* gene is vital in iron absorption as it acts by suppressing hepcidin, thereby allowing effective iron absorption [33]. Therefore, genetic variations

that lead to a loss-of-function of the genes that regulate iron absorption may predispose to low iron status.

Although the influence of nutritional deficiencies and inflammation/infection on anaemia has been widely investigated, the role of genetic risk factors has not received adequate attention, particularly in West Africa. Given the high burden of anaemia in Sub-Saharan Africa (SSA) and diverse genetic background of African populations [35,36], it is crucial to assess the impact of genetic risk factors on anaemia. This may aid in identifying better anaemia control measures.

2.3.1. Anaemia definition criteria

The haemoglobin (Hb) threshold is used to define anaemia. At the population level, measurement of Hb is the easiest and most reliable indication of anaemia [37]. The WHO set 12.0g/dL and 11.0g/dL as the threshold for defining anaemia in non-pregnant and pregnant women respectively [38]. However, these Hb cut-offs have been a subject of controversy, due to the argument that the thresholds may not be applicable universally [39]. This is attributed to the different population characteristics such as race, ethnicity and geographical location. For example, African-Americans have been found to have lower Hb levels compared to their White counterparts and these differences remained even after controlling for racial differences [40,41]. This demonstrates that applying the WHO criteria may increase the prevalence of anaemia in African-Americans. Similarly, people living in higher altitudes usually have higher Hb [42], therefore, applying the WHO threshold may lead to a lower prevalence of anaemia in such settings [38].

In sub-Saharan Africa, the presence of infectious diseases may also influence adaptation to low iron status [43]. For example, iron deficiency is considered to be a

protection against malaria [44], whereas, malaria is said to increases the risk of iron deficiency [45]. This has led to the hypothesis that the anaemia burden can be reduced by eliminating malaria [46]. This demonstrates the need to consider setting and population characteristics when determining anaemia prevalence using universal thresholds. Thus, it has been proposed to adjust for population characteristics when defining anaemia at the populations level [47], to avoid over-or under-estimation of anaemia prevalence in a given population.

2.4. Anaemia control strategies

Iron supplementation is the routine anaemia treatment and prevention approach [48]. This measure is implemented alone or alongside food fortification programmes targeted at vulnerable populations [21,48]. Oral iron supplementation is generally effective in anaemias caused by nutritional iron deficiency. Thus, the WHO recommends daily or intermitted iron supplementation in pregnant women and children under five years of age [49–51]. This policy guidline is advanced based of the evidence that intermittent iron supplementation has been associated with minimal side effects [52]. A comprehensive review by Penas-Rozas and colleagues found that both intermittent and daily iron supplementation has similar outcomes in both mothers and infants [52]. Despite decades of rolling out these strategies, the burden of anaemia remained high in low- and middle-income countries (LMICs), with Western Africa and East Asia carrying the highest burden [53]. Thus, reducing the burden of anaemia in the most affected populations remains a priority for governments and international organisations [54].

Therefore, in a quest to reduce the anaemia burden in the vulnerable groups, in 2012, the World Health Assembly resolved to take action to reduce by half the 2012 anaemia prevalence among women of reproductive age, by the year 2025 [55]. According to the World Health Organisation (WHO), one of the strategies to achieve this goal is to *improve identification, measurement and understanding of anaemia among women of reproductive age* [48,55]. Consequently, the WHO recommended the *implementation of evidence-based, setting and population-specific strategies, while taking into account the aetiology and prevalence of anaemia* [56]. However, implementing this tailored preventive and treatment strategies requires a clear understanding of the risk factors and significant drivers of anaemia. This is particularly critical in populations where the prevalence of anaemia remained high. Thus, it is crucial to gain a deeper understanding of the genetic determinants of anaemia particularly in settings characterised by a high burden of anaemia and disproportionately high genetic diversity, such as in African populations.

2.5. The need for better intervention methods

Anaemia continues to be a recalcitrant global health problem that has gained significant attention due to its consequences on human health and wellbeing [18]. The fight against anaemia is intense in LMICs because of the sustained high burden [57]. The most recent study on the global burden of anaemia [9] and related studies [37,58–60] reported high disparities in the prevalence of anaemia between different geographical locations and population. Also, there are within-country disparities among different populations. For example, in the US, the National Health and Nutrition Examination Surveys (NHANES) identified considerable differences in anaemia burden between racial groups, with black women having more anaemia compared to white women [60].

The search for better iron intervention strategies necessitates a deeper understanding of the drivers of anaemia. The ambitious target commissioned by World Health Assembly to reduce the 2012 anaemia burden in women of reproductive age by half by 2025 [61], brought the identification of the underlying genetic determinants of anaemia to the forefront. Identification of the genetic determinants of an ineffective response to iron supplementation and those that predispose individuals or populations to anaemia may help in developing population-specific or personalised interventions.

The discovery of hepcidin, the hormone that centrally regulates iron metabolism, brought new insights into the understanding of iron homeostasis [62–64]. Additionally, the recent explosion in genome-wide association studies (GWASs) enabled the discovery of numerous single nucleotide polymorphisms (SNPs) within the hepcidin-iron axis that are associated with impaired iron status [65–67]. However, most of the studies on the effects of SNPs within the hepcidin regulatory genes have been conducted in non-African populations.

There have been calls to increase human genetic research in the African continent [68,69]. The high genetic diversity among Africans motivated the interest to study genetics determinants of diseases affecting Africans. Results from human genetic studies conducted in Europeans are not readily transferrable to African populations for multiple reasons: 1) Europeans only carry a subset of the global human genetic diversity [69]; 2) There is variation between populations in their biological adaptations to infectious diseases [68]; 3) There are differences in allele frequencies of genetic variants and differences in linkage disequilibrium of alleles across global populations [36]. Thus, it is crucial to conduct genetic studies in African populations to determine the effects of common genetic variants on anaemia in Africans.

2.6. The role of hepcidin in iron homeostasis

The discovery of hepcidin as the master regulator of iron metabolism brought new insights into how body iron is regulated [70]. Hepcidin is a liver-synthesised 25-amino acid hormone, which is encoded by the *HAMP* (hepcidin antimicrobial peptide) gene [71,72]. Hepcidin acts by binding to ferroportin, the mammalian cellular iron transporter, thereby inducing ferroportin's internalisation and degradation [73]. This action of hepcidin regulates plasma iron levels by restricting iron transport from the gut and mobilisation from storage sites (stores and reticuloendothelial iron release) [8] (**Figure 2)**. Hepcidin deficiency leads to maximal levels of hight functional ferroportin, resulting in increased iron absorption, release and transport in the blood [74]. Conversely, hepcidin excess leads to decreased ferroportin availability, which results in diminished iron absorption [4,75] **Figure 2**. Hepcidin is synthesised in response to increased iron stores, high plasma iron concentration and inflammation [1]. Reduced extracellular iron, hypoxia and increased erythropoiesis, leads to diminished hepcidin transcription (**Figure 2)**. Conversely, elevated hepcidin levels prevent dietary iron absorption and recycling from macrophages, and mobilisation from hepatocytes, leading to reduced extracellular iron [76,77].

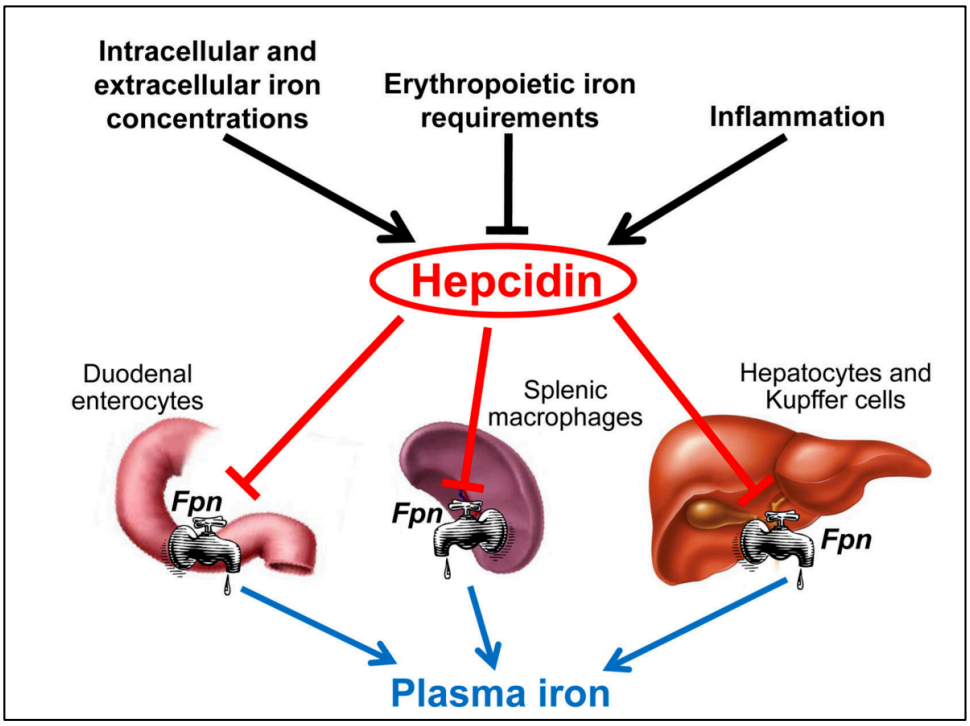


Figure 2. Hepcidin regulation of systemic iron distribution.

Adapted from Ganz and Nemeth 2012 [8].

The three major signals that influences hepcidin activity: Extracellular and intracellular iron, inflammation and erythropoiesis [8]. Hepcidin is elevated in response to inflammation, increased iron stores and high extracellular iron levels [1]. Conversely, increase erythropoietic drive lead to reduced hepcidin levels to allow increase iron mobilisation from the storage sites (liver and spleen), and absorption from the gut [76,77].

2.6.1. Hepcidin regulatory pathways

Hepcidin regulation of iron metabolism is mediated via three main molecular pathways, through the interconnection of genes and proteins [78]. The Janus associated kinase (JAK)/ signal transducer and activator of transcription 3 (STAT3) and bone morphogenetic protein (BMP)/ sons of mothers against decapentaplegic (SMAD) signalling pathways promote hepcidin synthesis [78]. In contrast, the hemochromatosis protein (HFE) – transferrin receptor 2 (TfR2) pathway suppresses hepcidin synthesis in response to elevate extracellular iron [79]. The JAK/STAT3 is activated in response to inflammatory stimuli, which induces interleukin 6 (IL6) production [80]. Increase

inflammation results in elevated IL6 which binds to its receptor (IL6R) thereby activating JAK1 [79]. Consequently, activated JAK1 causes the phosphorylation of STAT3 which moves to the nucleus to activate hepcidin production (**Figure 3A**).

The BMP/SMAD pathway is activated in response to increasing hepatic iron stores. Increased hepatic cellular iron induces BMP6 expression, which then interacts with BMPR and HJV, forming a complex [81]. The BMP6/BMPR interaction activates the SMAD pathway (**Figure 3B**). The SMAD pathway involves phosphorylation of regulatory SMAD1, 5 and 8 (PSMADs). The phosphorylated SMADs (pSMADs) complexed with SMAD4, and subsequent translocation of this complex to the nucleus results in the activation of hepcidin gene expression (**Figure 3B**). Furthermore, within the BMP/SMAD pathway, transmembrane protease serine 6 (TMPRSS6) (also referred to as matriptase 2) cleaves HJV to form a soluble HJV (sHJV), which inhibits BMP-induced hepcidin expression, as HJV is required for BMP/BMPR complex formation [81]. Thus, TMPRSS6 acts as a negative regulator of hepcidin through its interaction with HJV within the BMP/SMAD pathway. TMPRSS6 exerts its action in response to extracellular iron (**Figure 3B**). During increased iron stores, TMPRSS6 activity is reduced, allowing the BMP/SMAD pathway to continue which promotes hepcidin synthesis. Consequently, hepcidin elevation curtails iron absorption and release from storage sites [74]. Inversely, when extracellular iron levels are reduced, TMPRSS6 activity is increased to displace HJV, which in turn interferes with the BMP/SMAD pathway to suppress hepcidin transcription (**Figure 3B**). This action of TMPRSS6 on hepcidin promotes iron absorption from the enterocytes and release from storage sites [82].

The HFE/TfR2 pathway is activated when transferrin saturation increases and Tf-Fe^{2+} displaces HFE from TfR1 [83]. HFE then interacts with TfR2 to form the HFE/TfR2 complex [83]. Consequently, this complex activates hepcidin transcription via the HJV/BMP/SMAD and or extracellular signal-regulated kinase (ERK) – mitogen-associated protein kinase (MAPK) signalling pathways through a mechanism which is yet to be fully understood (**Figure 3C**) [84].

Hepcidin induction during iron overload reduces circulating plasma iron by restricting absorption and release of stored iron [74,76]. Similarly, hepcidin elevation during infection/inflammation promotes the reduction of iron availability to invading pathogens, a mechanism known as bacteriostatic hypoferremic response to infection/inflammation [8]. This mechanism is referred to as an evolutionary innate immune response to prevent worsening of infection [8]. Therefore, impaired hepcidin regulation has a vital role in modulating iron imbalance.

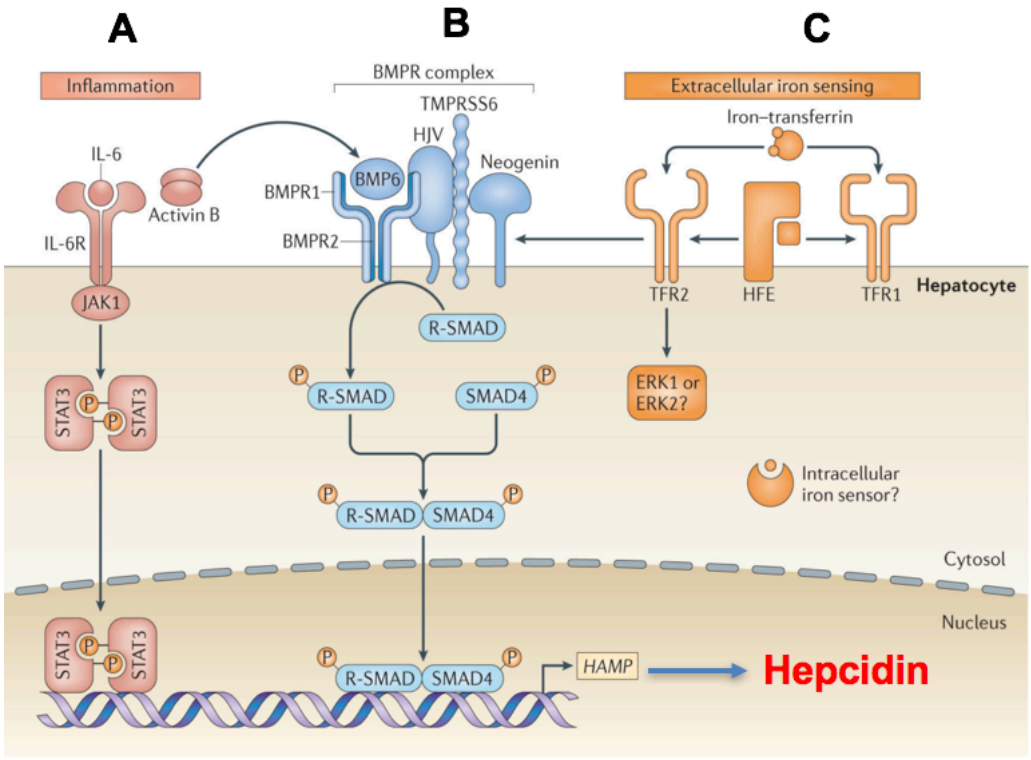


Figure 3. The molecular pathways governing hepcidin regulation of hepcidin transcription.

Adapted and modified from Ganz and Nemeth, 2015 [79].

Three main molecular pathways have been described to be involved in modulating hepcidin transcription: JAK/STAT3, BMP/SMAD and HFE/TfR2 signalling pathways [84]. In the inflammatory pathway (**A**), IL6 is induced in response to inflammatory stimuli, thereby activating the JAK/STAT3 pathway, which then triggers hepcidin production [84]. Similarly, in the BMP/SMAD pathway (**B**) increased hepatic cellular iron induces BMP6 expression, which then interacts with BMPR and HJV, forming a complex. TMPRSS6 and furin cleave HJV to form a soluble HJV (sHJV), which inhibits BMP-induced hepcidin expression, as HJV is required for BMP/BMPR complex formation [84]. The BMP/BMPR activates the SMAD pathway. The SMAD pathway involves phosphorylation of SMAD1, 5 and 8 (pSMADs). The formation of pSMADs/SMAD4 complex and subsequent translocation of this complex to the nucleus results in the activation of hepcidin gene expression [85,86]. (**C**) The HFE/TfR2 pathway is activated when transferrin saturation increases and Tf-Fe^{2+} displaces HFE from TfR1. HFE then interacts with TfR2 to form the HFE/TfR2 complex [83]. Consequently, this complex activates hepcidin transcription via the HJV/BMP/SMAD and or ERK/MAPK signalling pathways through a mechanism which is yet to be fully understood [84].

IL6, interleukin-6; IL6R, IL6 receptor; JAK, Janus associated kinase; STAT, signal transducer and activator of transcription; BMP, bone morphogenetic protein; BMPR, BMP receptor; TMPRSS6, transmembrane protease serine 6; HJV, hemojuvelin; SMAD, sons of mothers against decapentaplegic; pSMAD, phosphorylated SMAD; TfR, transferrin receptor; HFE, human haemochromatosis protein; ERK, extracellular signal-regulated kinase.

The hepcidin-iron relationship has been implicated in numerous iron-related pathologies, which are attributed to dysregulation of the hepcidin-iron axis [6]. Hepcidin excess is associated with anaemia due to inhibited iron absorption and decreased release from storage compartments [87]. Conversely, hepcidin suppression leads to haemochromatosis, characterised by the accumulation of excess iron in vital organs [4]. Haemochromatosis arises from decreased hepcidin synthesis, which promotes excess intestinal iron absorption, recycling and mobilisation of iron from old red cells and macrophages, thereby increasing extracellular iron [83].

Different stimuli promote either increase or decrease in hepcidin expression (**Figure 3).** Each of the two extremes can result in iron-related pathologies [88]. Altered hepcidin expression is associated with iron pathologies. Among the two principal iron pathologies considered to be mediated by impairment of the hepcidin regulatory genes, anaemia is the most common in sub-Saharan Africa (SSA) [10,22]. In iron deficiency anaemia (IDA), hepcidin is generally low, whereas, in anaemia of chronic disease (ACD) or inflammation, hepcidin is either normal or elevated [89]. These has led to the proposal that hepcidin levels can be used as a specific diagnostic marker to distinguish between IDA and ACD [90–92].

In addition to being a potential diagnostic marker, manipulation of hepcidin has been proposed to be a treatment of iron-related pathologies [93]. For example, hepcidin

antagonists have the potential to treat anaemias due to inappropriately elevated hepcidin [94]. Conversely, hepcidin agonists are proposed for use in treating diseases of loading anaemias [89]. Research into the usability of hepcidin as treatment of iron pathologies are currently under at various stages [93].

2.6.2. Transmembrane protease serine 6 (TMPRSS6) regulation of hepcidin

The role of TMPRSS6 (matriptase-2), a type 2 serine protease in regulating iron homeostasis was first discovered in mice which lacked this protease [95,96]. The mice presented with anaemia due to elevated levels of hepcidin and impaired intestinal iron absorption [95]. Further *in vitro* studies demonstrated that TMPRSS6 exerts its function by suppressing BMP6 stimulation of hepcidin transcription through cell surface proteolytic cleavage of the HJV, the BMP6 co-receptor [97]. Subsequently, human studies elucidated the role of TMPRSS6 in modulating iron homeostasis through its interaction with hepcidin [81]. Impaired TMPRSS6 function has been associated with iron deficiency and several single nucleotide polymorphisms (SNP) in the TMPRSS6 gene has been associated with iron-refractory iron deficiency (IRIDA) [98,33,99].

The *TMPRSS6* gene has 18 exons [66,96], and rs855791, rs2235321 and rs4820268 are the most commonly reported SNPs. The *TMPRSS6* rs8557891 is a non-synonymous SNPs on exon 17, which is caused by a G to A change leading to amino acid change from alanine to valine at the catalytic domain [66,96]. *TMPRSS6* rs855791 has been found to modulate hepcidin transcription *in vitro* [100]. *TMPRSS6* rs2235321 and rs4820268 are synonymous SNPs located on exon 13 and 17, respectively [33]. How these two synonymous SNPs affect *TMPRSS6* activity is unknown. Although non-coding variants and synonymous SNPs were previously not known to influence

phenotype, recent findings demonstrate that such variants may influence splicing or slow down the translation, and RNA transcription and regulation [101,102]. Therefore, given the role of TMPRSS6 in hepcidin regulation of iron homeostasis and its association with iron deficiency and impaired iron absorption, it is important to investigate the impact of the SNPs in this gene and those in other genes in the hepcidin regulatory pathway, in modulating low iron status in Africans.

2.7. Anaemia in the context of Gambia

The Gambia is situated at the coast of West Africa, and it is surrounded on three sides by Senegal, except on the Atlantic Ocean (**Figure 4**). The country has a population of 2.2 million as of October 2018, with a growth rate at 3.2% per year [103]. More than half of the population (approximately 63.6%) is below 25 years age and 52% of the population is between 15 and 59 years old [103].

Figure 4. A map of Africa showing the location of the Gambia [104].

The Gambia is classified as a low-income and food-deficit country [105], with tourism and agriculture as the principal foreign exchange-earners [106]. The predominant staple foods are rice and millet. Rice is regularly eaten in the urban areas, whereas millet is the main dish in most rural communities [105]. These foods are complemented with other cereals such as maize, sorhgum and Digitaria exilis (locally called Findi or Fonio) [105]. In urban areas, the typical lunch dish is rice, whilst bread is mainly consumed for dinner and breakfast [107]. All the main dishes are served with sauces made with oil or peanut butter paste and either fish or meat and other vegetables [105]. Meat consumption is generally low, and fish is the primary source of protein [108]. Recently,

there has been a massive increase in imported chicken, and this is increasingly consumed instead of fish or meat [105].

The Gambia is among the countries with the highest anaemia prevalence [58]. In 2012, a National Health Survey was conducted, which reported that more than 60% and 73% women of reproductive age and children under five years old, are anaemic respectively [108]. Also, a recent National Micronutrient Survey (MNS) reported that 50% of children under five years of age and non-pregnant women of reproductive age were anaemic [11]. From the MNS study, the prevalence of iron deficiency was 59% and 41% pre-school children and non-pregnant women of reproductive age respectively were iron deficient [11]. Although the prevalence of anaemia appears to decline, it remains a significant public health concern.

Due to the high anaemia prevalence, particularly among the most vulnerable populations, there have been various control programmes in place [108]. The anaemia control programmes include routine and intermittent iron supplementation for pregnant women and children [108]. Also, food fortification programmes are implemented, and children are regularly dewormed to prevent worm infestation [108]. The flour is the common staples that is fortified with vitamins [109], but new biofortified crops have been introduced in the country targeting rural communities [110]. These biofortified crops include vitamin A rich sweet potato and maize, and iron rich beans [111] and pearl millet [112].

Despite, decades of implementing anaemia control strategies particularly iron supplementation, the prevalence of anaemia remains high, particularly among the vulnerable groups (women of reproductive age and pre-school children). Therefore, it is crucial to investigate the underlying causes of anaemia in The Gambia and in similar settings to improve future iron intervention strategies.

2.8. The need to investigate the genetic influences of anaemia

The discovery of hepcidin and its related regulatory proteins renewed the interest in understanding the genetic influences of iron related pathologies [4,78,113]. Dysregulation of the *TMPRSS6* genes is one of the most commonly studied genetic determinants of anaemia, due to its association with iron-refractory iron deficiency anaemia (IRIDA) [66,114,115]. IRIDA has been described mainly in non-African populations [33,66,116]. However, given the high prevalence of anaemia in Africans, it is relevant to investigate the influence of these known genetic variants associated with anaemia in African populations.

Also, apart from *TMPRSS6*, defects in other iron regulatory genes such as *SCL11A2*, *SLC40A1* and transferrin (*TF*) have been linked to impaired iron status [4,117]. *SLC11A2* encodes the divalent metal transporter 1 (DMT1) protein which is involved in iron transport across the duodenum in the duodenal enterocytes [32,118]. In mice models, impairment of *SLC11A2* has been associated with microcytic anaemia due to impaired iron absorption [119,120]. Subsequently, human studies revealed the association between *SLC11A2* variants and microcytic anaemia [32,121]. Similarly, defects in *SLC40A1* (the gene that encodes ferroportin) is associated with increased hepatic iron accumulation, accompanied by microcytic anaemia in Africans [122,123]. This happens due the inability of hepcidin to downregulate ferroportin, which leads to inappropriate iron transport [88]. Also, defects in the transferrin gene (*TF*) has been associated with hypotranferrinaemia, which leads to iron-loading [124]. Therefore, it is possible that genetic variants in genes involved in iron regulation could cause a range of degrees of iron deficiency and anaemia.

The majority of the early studies on genetic influences of iron status were conducted in European populations [125–127]. Most of these studies focused on identifying the

genetic influences of iron overload syndromes [128,129], including studies on hereditary hemochromatosis (HH). HH is a rare disorder that mainly affects populations of European ancestry and primarily caused by a rare genetic defect in the *HFE* gene [130]. Recently, there has been an increased interest in understanding the genetic influences of low iron status. Several studies have observed associations between genetic variants and iron status biomarkers [131–133]. Consistently, studies have implicated *TMPRSS6* (the gene that encodes matriptase-2) as one of the genes that modulate iron status [100,134,135]. Numerous SNPs in the *TMPRSS6* gene have been linked to abnormal iron biomarkers and haematological traits, including serum iron concentrations, transferrin saturation, haemoglobin, erythrocyte count and mean corpuscular volume [82,136–139].

Despite the growing interest in understanding genetic influences of anaemia and low iron status, there is a scarcity of data in this regard on Africans populations. This is particularly essential given that African populations have disproportionately high genetic diversity. Furthermore, it is particularly incredible to conduct genetic studies of anaemia and iron deficiency in African populations to examine whether the genetic effects observed in Europeans and Asian are replicable. Also, given the environmental factors that favour anaemia in Africans, including diet with low bioavailable iron, infectious diseases (e.g. malaria), it is necessary to determine the impact of host genetic factors in modulating iron status. Understanding the genetic determinants of anaemia and iron deficiency in African populations may help to improve future iron intervention strategies.

2.9. Importance of the MRCG Keneba Biobank

The Keneba Biobank at MRCG at LSHTM is a collection of biological samples and phenotype data from the residents of the Kiang West district, located in the Lower River Region of The Gambia [140], **Figure 5**. This district has a population of approximately 16,000 people distributed among 36 villages. The Kiang West Demographic Surveillance System (KWDSS) at the MRCG Keneba was established in 2004. Since then, the KWDSS has been monitoring and recording migration and vital statistics (e.g. birth, deaths) for the entire population of this district.

The Biobank was established in 2012, and it utilises the KWDSS to recruit study participants. The Keneba Biobank has already recruited 11,5000 individuals. From these, 3116 individuals have been genotyped using the Illumina Human Exome array, providing data on 80K SNPs. The KWDSS, in conjunction with the Keneba Biobank, offers a unique platform for studies involving the whole of the rural population of Kiang West, and enables 'recall' of study participants for genetic studies.

Figure 5. Map of the Gambia and Kiang West showing the study location. Adapted and modified from Hennig et al. 2017 [104,140].

2.10. Aims and Objectives

The aim of this PhD was to investigate the role of genetic variants in the human hepcidin and iron regulatory genes on the risk anaemia and impaired response to oral supplementation in Africans.

The objectives of this PhD are:

a) To identify the genetic variants within the iron and hepcidin regulatory genes that influence iron status in global populations.
b) To investigate the effects of common genetic variants in the iron regulatory genes on iron status in health rural Gambians.
c) To examine the impact of *TMPRSS6* SNPs on oral iron absorption in healthy population from rural Gambia using a recall-by-genotype approach.

2.11. References

1. Ganz, T. Systemic Iron Homeostasis. *Physiological Reviews* **93**, 1721–1741 (2013).
2. Oliveira, F., Rocha, S. & Fernandes, R. Iron Metabolism: From Health to Disease: Iron Metabolism. *J. Clin. Lab. Anal.* **28**, 210–218 (2014).
3. Andrews, N. C. Iron metabolism: iron deficiency and iron overload. *Annu. Rev. Genom. Hum. Genet.* **1**, 75–98 (2000).
4. Silva, B. & Faustino, P. An overview of molecular basis of iron metabolism regulation and the associated pathologies. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1852**, 1347–1359 (2015).
5. Michels, K., Nemeth, E., Ganz, T. & Mehrad, B. Hepcidin and Host Defense against Infectious Diseases. *PLoS Pathog* **11**, e1004998 (2015).
6. Drakesmith, H. & Prentice, A. M. Hepcidin and the Iron-Infection Axis. *Science* **338**, 768–772 (2012).
7. Gozzelino, R. & Arosio, P. Iron Homeostasis in Health and Disease. *IJMS* **17**, 130 (2016).
8. Ganz, T. & Nemeth, E. Hepcidin and iron homeostasis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 1434–1443 (2012).
9. Kassebaum, N. J. The Global Burden of Anemia. *Hematology/Oncology Clinics of North America* **30**, 247–308 (2016).
10. Pasricha, S.-R. Anemia: a comprehensive global estimate. *Blood* **123**, 611–612 (2014).
11. Petry, N. *et al.* Micronutrient Deficiencies, Nutritional Status and the Determinants of Anemia in Children 0–59 Months of Age and Non-Pregnant Women of Reproductive Age in The Gambia. *Nutrients* **11**, 2275 (2019).

12. Pollitt, E. Iron Deficiency and Cognitive Function. *Annu. Rev. Nutr.* **13**, 521–37 (1993).
13. Pivina, L., Semenova, Y., Doşa, M. D., Dauletyarova, M. & Bjørklund, G. Iron Deficiency, Cognitive Functions, and Neurobehavioral Disorders in Children. *J Mol Neurosci* **68**, 1–10 (2019).
14. Balarajan, Y., Ramakrishnan, U., Özaltin, E., Shankar, A. H. & Subramanian, S. Anaemia in low-income and middle-income countries. *The Lancet* **378**, 2123–2135 (2011).
15. Jáuregui-Lobera, I. Iron deficiency and cognitive functions. *NDT* 2087 (2014) doi:10.2147/NDT.S72491.
16. Stauder, R. & Thein, S. L. Anemia in the elderly: clinical implications and new therapeutic concepts. *Haematologica* **99**, 1127–1130 (2014).
17. World Health Organization. the Global Prevalence of Anaemia in 2011. *WHO Report* 48 (2011) doi:10.1017/S1368980008002401.
18. Kassebaum, N. J. The Global Burden of Anemia. *Hematology/Oncology Clinics of North America* **30**, 247–308 (2016).
19. World Health Organization. *THE GLOBAL PREVALENCE OF ANAEMIA IN 2011*. (2015).
20. Petry, N. *et al.* The Proportion of Anemia Associated with Iron Deficiency in Low, Medium, and High Human Development Index Countries: A Systematic Analysis of National Surveys. *Nutrients* **8**, 693 (2016).
21. World Health Organization. *Essential nutrition actions: improving maternal, newborn, infant and young child health and nutrition.* (World Health Organization, 2013).

22. Pasricha, S.-R. & Drakesmith, H. Iron Deficiency Anemia. *Hematology/Oncology Clinics of North America* **30**, 309–325 (2016).
23. Kassebaum, N. J. *et al.* A systematic analysis of global anemia burden from 1990 to 2010. *Blood* **123**, 615–624 (2014).
24. Zimmermann, M. B., Chaouki, N. & Hurrell, R. F. Iron deficiency due to consumption of a habitual diet low in bioavailable iron: a longitudinal cohort study in Moroccan children. *The American Journal of Clinical Nutrition* **81**, 115–121 (2005).
25. Mwangi, M. *et al.* Iron for Africa—Report of an Expert Workshop. *Nutrients* **9**, 576 (2017).
26. Samtiya, M., Aluko, R. E. & Dhewa, T. Plant food anti-nutritional factors and their reduction strategies: an overview. *Food Prod Process and Nutr* **2**, 6 (2020).
27. Beard, J. L. Iron Requirements in Adolescent Females. *The Journal of Nutrition* **130**, 440S-442S (2000).
28. WHO. Soil-transmitted helminth infections. https://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections (2020).
29. White, N. J. Anaemia and malaria. *Malar J* **17**, 371 (2018).
30. Prentice, A. M. *et al.* Respiratory infections drive hepcidin-mediated blockade of iron absorption leading to iron deficiency anemia in African children. *Sci. Adv.* **5**, eaav9020 (2019).
31. Kohne, E. Hemoglobinopathies. *Deutsches Aerzteblatt Online* (2011) doi:10.3238/arztebl.2011.0532.
32. Mims, M. P. *et al.* Identification of a human mutation of DMT1 in a patient with microcytic anemia and iron overload. *Blood* **105**, 1337–1342 (2005).

33. Wang, C.-Y., Meynard, D. & Lin, H. Y. The role of TMPRSS6/matriptase-2 in iron regulation and anemia. *Front. Pharmacol.* **5**, (2014).
34. Gulec, S., Anderson, G. J. & Collins, J. F. Mechanistic and regulatory aspects of intestinal iron absorption. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **307**, G397–G409 (2014).
35. Tishkoff, S. A. *et al.* The Genetic Structure and History of Africans and African Americans. *Science* **324**, 1035–1044 (2009).
36. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26–31 (2019).
37. World Health Organization. *Worldwide prevalence of anaemia 1993–2005: WHO Global Database on Anaemia*. https://apps.who.int/iris/bitstream/handle/10665/43894/97892?sequence=1 (2008).
38. WHO. *Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Vitamin and Mineral Nutrition Information System, WHO/NMH/NHD/MNM/11.1)*. (2011).
39. Schechter, G. P. Hemoglobin levels in African-Americans. *Blood* **107**, 2208–2209 (2006).
40. Zakai, N. A. *et al.* Correlates of Anemia in American Blacks and Whites: The REGARDS Renal Ancillary Study. *American Journal of Epidemiology* **169**, 355–364 (2008).
41. Williams, D. M. Racial differences of hemoglobin concentration: measurements of iron, copper, and zinc. *The American Journal of Clinical Nutrition* **34**, 1694–1700 (1981).

42. Silubonde, T. M. *et al.* Adjusting Haemoglobin Values for Altitude Maximizes Combined Sensitivity and Specificity to Detect Iron Deficiency among Women of Reproductive Age in Johannesburg, South Africa. *Nutrients* **12**, 633 (2020).
43. Jonker, F. A. M., te Poel, E., Bates, I. & Boele van Hensbroek, M. Anaemia, iron deficiency and susceptibility to infection in children in sub-Saharan Africa, guideline dilemmas. *Br J Haematol* **177**, 878–883 (2017).
44. Gwamaka, M. *et al.* Iron Deficiency Protects Against Severe Plasmodium falciparum Malaria and Death in Young Children. *Clinical Infectious Diseases* **54**, 1137–1144 (2012).
45. Spottiswoode, N., Duffy, P. E. & Drakesmith, H. Iron, anemia and hepcidin in malaria. *Front. Pharmacol.* **5**, (2014).
46. Muriuki, J. & Atkinson, S. How Eliminating Malaria May Also Prevent Iron Deficiency in African Children. *Pharmaceuticals* **11**, 96 (2018).
47. Wirth, J. P. *et al.* Predictors of anemia in women of reproductive age: Biomarkers Reflecting Inflammation and Nutritional Determinants of Anemia (BRINDA) project. *Am J Clin Nutr* **106**, 416S-427S (2017).
48. World Health Organisation. *Strategies to prevent anaemia: Recommendations from an Expert Group Consultation*. 35 http://origin.searo.who.int/entity/nutrition/recommendations_on_anaemia1.pdf (2016).
49. Pasricha, S.-R., Drakesmith, H., Black, J., Hipgrave, D. & Biggs, B.-A. Control of iron deficiency anemia in low- and middle-income countries. *Blood* **121**, 2607–2617 (2013).

50. McGuire, S. World Health Organization. Comprehensive Implementation Plan on Maternal, Infant, and Young Child Nutrition. Geneva, Switzerland, 2014. *Advances in Nutrition* **6**, 134–135 (2015).
51. Suchdev, P. S., Peña-Rosas, J. P. & De-Regil, L. M. Multiple micronutrient powders for home (point-of-use) fortification of foods in pregnant women. *Cochrane Database of Systematic Reviews* (2015) doi:10.1002/14651858.CD011158.pub2.
52. Peña-Rosas, J. P., De-Regil, L. M., Gomez Malave, H., Flores-Urrutia, M. C. & Dowswell, T. Intermittent oral iron supplementation during pregnancy. *Cochrane Database of Systematic Reviews* (2015) doi:10.1002/14651858.CD009997.pub2.
53. Pasricha, S.-R., Hayes, E., Kalumba, K. & Biggs, B.-A. Effect of daily iron supplementation on health in children aged 4–23 months: a systematic review and meta-analysis of randomised controlled trials. *The Lancet Global Health* **1**, e77–e86 (2013).
54. World Health Organization. Nutritional Anaemias: Tools for Effective Prevention and Control. (2017).
55. World Health Organisation. *WHA Global Nutrition Targets 2020: Anaemia Policy Brief*. https://www.who.int/nutrition/topics/globaltargets_anaemia_policybrief.pdf?ua=1- (2014).
56. World Health Organisation. *Joint statement by the World Health Organization and the United Nations Children's Fund: Focusing on anaemia*. (2004).
57. World Health Organization(WHO). *Essential Nutrition Actions: Improving Maternal, Newborn, Infant and Young Child Health and Nutrition*. (World Health Organisation Press, 2013).

58. Stevens, G. A. *et al.* Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995–2011: a systematic analysis of population-representative data. *The Lancet Global Health* **1**, e16–e25 (2013).
59. Hu, S. *et al.* Disparity of anemia prevalence and associated factors among rural to urban migrant and the local children under two years old: a population based cross-sectional study in Pinghu, China. *BMC Public Health* **14**, 601 (2014).
60. Le, C. H. H. The Prevalence of Anemia and Moderate-Severe Anemia in the US Population (NHANES 2003-2012). *PLoS ONE* **11**, e0166635 (2016).
61. World Health Organization. Global Nutrition targets 2015 Anaemia Policy Brief. *Global Nutrition Targets 2025* **2**, 8 (2014).
62. Ganz, T. Systemic iron homeostasis. 1721–1741 (2013) doi:10.1152/physrev.00008.2013.
63. Silva, B. & Faustino, P. An overview of molecular basis of iron metabolism regulation and the associated pathologies. *Biochimica et biophysica acta* **1852**, 1347–1359 (2015).
64. Collins, J. F., Wessling-Resnick, M. & Knutson, M. D. Hepcidin regulation of iron transport. *The Journal of nutrition* **138**, 2284–2288 (2008).
65. Tanaka, T. *et al.* A genome-wide association analysis of serum iron concentrations. *Blood* **115**, 94–96 (2010).
66. Lee, P. Role of Matriptase-2 (TMPRSS6) in Iron Metabolism. *Acta Haematol* **122**, 87–96 (2009).
67. Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet* **41**, 1182–1190 (2009).

68. Gomez, F., Hirbo, J. & Tishkoff, S. A. Genetic Variation and Adaptation in Africa: Implications for Human Evolution and Disease. *Cold Spring Harbor Perspectives in Biology* **6**, a008524–a008524 (2014).
69. Campbell, M. C. & Tishkoff, S. A. African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu. Rev. Genom. Hum. Genet.* **9**, 403–433 (2008).
70. Ganz, T. Hepcidin and iron regulation, 10 years later. *Blood* **117**, 4425–4433 (2011).
71. Krause, A. *et al.* LEAP-1, a novel highly disulfide-bonded human peptide, exhibits antimicrobial activity. *FEBS Letters* **480**, 147–150 (2000).
72. Park, C. H., Valore, E. V., Waring, A. J. & Ganz, T. Hepcidin, a Urinary Antimicrobial Peptide Synthesized in the Liver. *J. Biol. Chem.* **276**, 7806–7810 (2001).
73. Nemeth, E. *et al.* Hepcidin Regulates Cellular Iron Efflux by Binding to Ferroportin and Inducing Its Internalization. *Science* **306**, 2090–2093 (2004).
74. Zhao, N., Zhang, A.-S. & Enns, C. A. Iron regulation by hepcidin. *J. Clin. Invest.* **123**, 2337–2343 (2013).
75. Rishi, G., Wallace, D. F. & Subramaniam, V. N. Hepcidin: regulation of the master iron regulator. *Bioscience Reports* **35**, e00192 (2015).
76. Ruchala, P. & Nemeth, E. The pathophysiology and pharmacology of hepcidin. *Trends in Pharmacological Sciences* **35**, 155–161 (2014).
77. Girelli, D., Nemeth, E. & Swinkels, D. W. Hepcidin in the diagnosis of iron disorders. *Blood* **127**, 2809–2813 (2016).

78. Mleczko-Sanecka, K. *et al.* Unbiased RNAi screen for hepcidin regulators links hepcidin suppression to proliferative Ras/RAF and nutrient-dependent mTOR signaling. *Blood* **123**, 1574–1585 (2014).
79. Ganz, T. & Nemeth, E. Iron homeostasis in host defence and inflammation. *Nat Rev Immunol* **15**, 500–510 (2015).
80. Verga Falzacappa, M. V. *et al.* STAT3 mediates hepatic hepcidin expression and its inflammatory stimulation. *Blood* **109**, 353–358 (2007).
81. Finberg, K. E., Whittlesey, R. L., Fleming, M. D. & Andrews, N. C. Down-regulation of Bmp/Smad signaling by Tmprss6 is required for maintenance of systemic iron homeostasis. *Blood* **115**, 3817–3826 (2010).
82. Ramsay, A. J., Hooper, J. D., Folgueras, A. R., Velasco, G. & Lopez-Otin, C. Matriptase-2 (TMPRSS6): a proteolytic regulator of iron homeostasis. *Haematologica* **94**, 840–849 (2009).
83. D'Alessio, F., Hentze, M. W. & Muckenthaler, M. U. The hemochromatosis proteins HFE, TfR2, and HJV form a membrane-associated protein complex for hepcidin regulation. *Journal of Hepatology* **57**, 1052–1060 (2012).
84. Kong, W.-N., Gao, G. & Chang, Y.-Z. Hepcidin and sports anemia. *Cell Biosci* **4**, 19 (2014).
85. Wang, C.-Y. & Babitt, J. L. Hepcidin regulation in the anemia of inflammation: *Current Opinion in Hematology* **23**, 189–197 (2016).
86. Sangkhae, V. & Nemeth, E. Regulation of the Iron Homeostatic Hormone Hepcidin. *Adv Nutr* **8**, 126–136 (2017).
87. Meynard, D., Babitt, J. L. & Lin, H. Y. The liver: conductor of systemic iron balance. *Blood* **123**, 168–176 (2014).

88. Camaschella, C. & Pagani, A. Advances in understanding iron metabolism and its crosstalk with erythropoiesis. *Br J Haematol* **182**, 481–494 (2018).
89. Pagani, A., Nai, A., Silvestri, L. & Camaschella, C. Hepcidin and Anemia: A Tight Relationship. *Front. Physiol.* **10**, 1294 (2019).
90. Rochette, L. *et al.* The iron-regulatory hormone hepcidin: A possible therapeutic target? *Pharmacology & Therapeutics* **146**, 35–52 (2015).
91. Sebastiani, G., Wilkinson, N. & Pantopoulos, K. Pharmacological Targeting of the Hepcidin/Ferroportin Axis. *Front. Pharmacol.* **7**, (2016).
92. Vyoral, D. & Jiri Petrak. Therapeutic potential of hepcidin − the master regulator of iron metabolism. *Pharmacological Research* **115**, 242–254 (2017).
93. Katsarou, A. & Pantopoulos, K. Hepcidin Therapeutics. *Pharmaceuticals* **11**, 127 (2018).
94. Poli, M., Asperti, M., Ruzzenenti, P., Regoni, M. & Arosio, P. Hepcidin antagonists for potential treatments of disorders with hepcidin excess. *Front. Pharmacol.* **5**, (2014).
95. Folgueras, A. R. *et al.* Matriptase-2 deficiency protects from obesity by modulating iron homeostasis. *Nat Commun* **9**, 1350 (2018).
96. Du, X. *et al.* The Serine Protease TMPRSS6 Is Required to Sense Iron Deficiency. *Science* **320**, 1088–1092 (2008).
97. Silvestri, L. *et al.* The Serine Protease Matriptase-2 (TMPRSS6) Inhibits Hepcidin Activation by Cleaving Membrane Hemojuvelin. *Cell Metabolism* **8**, 502–511 (2008).
98. Finberg, K. E. *et al.* Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA). *Nat Genet* **40**, 569–571 (2008).

99. Sato, T. *et al.* Novel missense mutation in the TMPRSS6 gene in a Japanese female with iron-refractory iron deficiency anemia. *Int J Hematol* **94**, 101–103 (2011).

100. Nai, A. *et al.* TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. *Blood* **118**, 4459–4462 (2011).

101. Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* **12**, 683–691 (2011).

102. Wang, G.-S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8**, 749–761 (2007).

103. Access Gambia. Population Figures For Gambia. http://www.accessgambia.com/information/population.html (2020).

104. worldatlas.com. Geography Statistics Of Gambia. https://www.worldatlas.com/webimage/countrys/africa/gambia/gmlandst.htm (2020).

105. FAO. *NUTRITION COUNTRY PROFILE REPUBLIC OF THE GAMBIA 2010*. (2010).

106. World Bank Group. The World Bank In The Gambia. https://www.worldbank.org/en/country/gambia/overview#1 (2020).

107. Access Gambia. Cooking Recipes in Gambia. https://www.accessgambia.com/information/food-recipes.html#:~:text=For%20the%20majority%20of%20Gambians,spices%20and%20sometimes%20peanut%20butter. (2020).

108. Gambia Bureau of Statistics Banjul, The Gambia. *The Gambia Demographic and Health Survey 2013*. (2014).

109. NESSIM Trading Company Ltd. Wheat flour. https://nessimtrading.com/our-products/.

110. THE GAMBIA – EU COOPERATION. Improving Food Security and Nutrition in The Gambia through Food Fortification. http://www.gambia-ec.gm/news/food-fortification-under-the-project-improving-food-security-and-nutrition-in-the-gambia-through-food-fortification-implemented-by-fao/ (2018).

111. FAO Gambia. Improving food security and nutrition in The Gambia through food fortification. (2018).

112. Herrington, C., Lividini, K., Angel, M. D. & Birol, E. *Prioritizing Countries for Biofortification Interventions: Biofortification Priority Index Second Edition (BPI 2.0)*. https://www.harvestplus.org/content/prioritizing-countries-biofortification-interventions-biofortification-priority-index-second (2019).

113. Parrow, N. L. & Fleming, R. E. Bone Morphogenetic Proteins as Regulators of Iron Metabolism. *Annu. Rev. Nutr.* **34**, 77–94 (2014).

114. Lakhal, S. *et al.* Regulation of type II transmembrane serine proteinase TMPRSS6 by hypoxia-inducible factors: new link between hypoxia signaling and iron homeostasis. *J. Biol. Chem.* **286**, 4090–4097 (2011).

115. De Falco, L. *et al.* Functional and clinical impact of novel TMPRSS6 variants in iron-refractory iron-deficiency anemia patients and genotype-phenotype studies. *Human Mutation* n/a-n/a (2014) doi:10.1002/humu.22632.

116. Pei, S.-N. *et al.* TMPRSS6 rs855791 Polymorphism Influences the Susceptibility to Iron Deficiency Anemia in Women at Reproductive Age. *Int. J. Med. Sci.* **11**, 614–619 (2014).

117. Donker, A. E., Brons, P. P. T. & Swinkels, D. W. Microcytic anaemia with low transferrin saturation, increased serum hepcidin and non-synonymous *TMPRSS6*

variants: not always iron-refractory iron deficiency anaemia. *Br J Haematol* **169**, 150–151 (2015).

118. Morgan, E. H. & Oates, P. S. Mechanisms and Regulation of Intestinal Iron Absorption. *Blood Cells, Molecules, and Diseases* **29**, 384–399 (2002).

119. Gunshin, H. *et al.* Slc11a2 is required for intestinal iron absorption and erythropoiesis but dispensable in placenta and liver. *J. Clin. Invest.* **115**, 1258–1266 (2005).

120. De Falco, L. *et al.* Identification and characterization of the first SLC11A2 isoform 1a mutation causing a defect in splicing process and an hypomorphic allele expression of the *SLC11A2* gene. *Br J Haematol* **159**, 492–495 (2012).

121. Beaumont, C. Two new human DMT1 gene mutations in a patient with microcytic anemia, low ferritinemia, and liver iron overload. *Blood* **107**, 4168–4170 (2006).

122. Gordeuk, V. R. *et al.* Iron overload in Africans and African-Americans and a common mutation in the SCL40A1 (ferroportin 1) gene. *Blood Cells, Molecules, and Diseases* **31**, 299–304 (2003).

123. Kasvosve, I. *et al.* Effect of ferroportin Q248H polymorphism on iron status in African children. *The American Journal of Clinical Nutrition* **82**, 1102–1106 (2005).

124. Donker, A. E. *et al.* Practice guidelines for the diagnosis and management of microcytic anemias due to genetic disorders of iron metabolism or heme synthesis. *Blood* **123**, 3873–3886 (2014).

125. Lim, E. M., Rossi, E., De Boer, W. B., Reed, W. D. & Jeffrey, G. P. Hepatic iron loading in patients with compound heterozygous HFE mutations. *Liver Int* **24**, 631–636 (2004).

126. Barbosa, K. V. B. D. *et al.* Hereditary Hemochromatosis: Population Screening Based on Phenotype in Brazilian Blood Donors. *Journal of Clinical Gastroenterology* **39**, 430–434 (2005).

127. Agudo, A. *et al.* Hemochromatosis (HFE) gene mutations and risk of gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Carcinogenesis* **34**, 1244–1250 (2013).

128. Bozzini, C. *et al.* Biochemical and genetic markers of iron status and the risk of coronary artery disease: an angiography-based study. *Clin Chem* **48**, 622–628 (2002).

129. Njajou, O. T. *et al.* A population-based study of the effect of the HFE C282Y and H63D mutations on iron metabolism. *Eur J Hum Genet* **11**, 225–231 (2003).

130. Adams, P. C. *et al.* Hemochromatosis and Iron-Overload Screening in a Racially Diverse Population. *N Engl J Med* **352**, 1769–1778 (2005).

131. Ganesh, S. K. *et al.* Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* **41**, 1191–1198 (2009).

132. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* **42**, 210–215 (2010).

133. Lo, K. S. *et al.* Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum Genet* **129**, 307–317 (2011).

134. Benyamin, B. *et al.* Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. *Nat Genet* **41**, 1173–1175 (2009).

135. McLaren, C. E. *et al.* Genome-Wide Association Study Identifies Genetic Loci Associated with Iron Deficiency. *PLoS ONE* **6**, e17390 (2011).

136. Kullo, I. J., Ding, K., Jouni, H., Smith, C. Y. & Chute, C. G. A Genome-Wide Association Study of Red Blood Cell Traits Using the Electronic Medical Record. *PLoS ONE* **5**, e13011 (2010).

137. Traglia, M. *et al.* Association of HFE and TMPRSS6 genetic variants with iron and erythrocyte parameters is only in part dependent on serum hepcidin concentrations. *Journal of Medical Genetics* **48**, 629–634 (2011).

138. Valenti, L. *et al.* The A736V TMPRSS6 Polymorphism Influences Hepatic Iron Overload in Nonalcoholic Fatty Liver Disease. *PLoS ONE* **7**, e48804 (2012).

139. Yaish, H. M. *et al.* Two novel mutations in TMPRSS6 associated with iron-refractory iron deficiency anemia in a mother and child. *Blood Cells, Molecules, and Diseases* **65**, 38–40 (2017).

140. Hennig, B. J. *et al.* Cohort Profile: The Kiang West Longitudinal Population Study (KWLPS)—a platform for integrated research and health care provision in rural Gambia. *Int. J. Epidemiol.* dyv206 (2017) doi:10.1093/ije/dyv206.

Chapter 3:

Differences in allele frequencies of genetic variants associated with iron imbalance among global populations

Chapter description:

This chapter presents the results of narative review paper and the assessment of the differences in allele frequencies of single nucleotide polymorphisms (SNP) associated with iron imbalance across global populations. This research paper has been published in PLOS ONE. https://dx.plos.org/10.1371/journal.pone.0235141

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1513421	**Title**	Mr.
First Name(s)	Momodou W.		
Surname/Family Name	Jallow		
Thesis Title	The impact of single nucleotide polymorphisms in human genes that regulate hepcidin and iron on oral iron absorption and the risk of anaemia in Africans		
Primary Supervisor	Dr Susana Campino		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	**PLOS ONE**		
When was the work published?	**July 2020**		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	**This work was done after registration as part of the thesis**		
Have you retained the copyright for the work?*	**NO**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be Published?	**N/A**
Please list the paper's authors in the intended authorship order:	**N/A**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary) **SECTION E**	**I am the main corresponding author. I contributed to conceptualisation and design of this study. I did the literature review and data analysis. I drafted the manuscript, managed co-author comments, and the submission process**

Student Signature	[redacted]
Date	**13 October 2020**

Supervisor Signature	[redacted]
Date	**13 October 2020**

Differences in the frequency of genetic variants associated with iron imbalance among global populations

Momodou W. Jallow[1,2], Carla Cerami[1], Taane G. Clarke[2], Andrew M. Prentice[1] and Susana Campino[2]

1 Nutrition Theme, MRC Unit The Gambia at London School of Hygiene & Tropical Medicine, Atlantic Road Fajara, P.O. Box 273, Banjul the Gambia

2 Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT.

Corresponding author: Momodou W. Jallow (mwjallow@mrc.gm)

Published in PLOS ONE: https://doi.org/10.1371/journal.pone.0235141

Abstract

Iron deficiency anaemia is a major health problem affecting approximately 1.2 billion people worldwide. Young children, women of reproductive age and pregnant women living in sub-Saharan Africa are the most vulnerable. It is estimated that iron deficiency accounts for half of anaemia cases. Apart from nutritional deficiency, infection, inflammation and genetic factors are the major drivers of anaemia. However, the role of genetic risk factors has not been thoroughly investigated. This is particularly relevant in African populations, as they carry high genetic diversity and have a high prevalence of anaemia. Multiple genetic variations in iron regulatory genes have been linked to impaired iron status. Here we conducted a literature review to identify genetic variants associated with iron imbalance among global populations. We compare their allele frequencies and risk scores and we investigated population-specific selection among populations of varying geographic origin using data from the Keneba Biobank representing individuals in rural Gambia and the 1000 Genomes Project. We identified a significant lack of data on the genetic determinants of iron status in sub-Saharan Africa. Most of the studies on genetic determinants of iron status have been conducted in Europeans. Also, we identified population differences in allele frequencies in candidate putative genetic risk factors. Given the disproportionately high genetic diversity in African populations coupled with their high prevalence of iron deficiency, there is need to investigate the genetic influences of low iron status in Sub-Saharan Africa. The resulting insights may inform the future implementation of iron intervention strategies.

Key words: Anaemia, Iron deficiency, iron, Genetic variants, iron imbalances, African populations.

Introduction

Iron deficiency anaemia (IDA) is a major health problem affecting approximately 1.2 billion people worldwide [1]. It was estimated to account for the 7th leading cause of disability worldwide in 2017 [2]. IDA is regarded as the dominant cause of anaemia, accounting for approximately 60% of the global anaemia burden[3]. Pre-school children and women of childbearing age in low- and middle-income countries are the most vulnerable [3,4], particularly those living in sub-Saharan Africa, where anaemia prevalence in the general population exceeds 40% [3]. This high prevalence of IDA persists despite the existence of aggressive iron supplementation programmes for vulnerable populations (women of childbearing age and children) [5–7].

Although iron supplementation can be effective in nutritional IDA, it is ineffective in non-nutritional IDA, particularly those caused by genetic factors [8]. Therefore, the identification of the major drivers of IDA in sub-Saharan Africa is required to inform new strategies. The discovery of hepcidin and other proteins involved in iron regulation have led to the identification of genetic factors associated with altered iron homeostasis [9–11]. Several genetic variants within the iron regulatory genes have been associated with imbalances in iron homeostasis, which could lead either to iron deficiency or overload [12–16]. Genetic variants leading to excess body iron occur mainly in the haemochromatosis (*HFE*) gene but are also seen in hepcidin (hepcidin antimicrobial peptide (*Hamp*)), transferrin receptor 2 (*TFR2*), solute carrier family 40 member 1 (*SLC40A1*), haemojuvelin (*HJV*) and transferrin (*TF*) genes [9–11]. These loci have important functions in the iron homeostasis pathways. For example, hepcidin regulates iron absorption and release [17]. Genetic polymorphisms in genes involved in the hepcidin suppressive pathway such as *TMPRSS6* (transmembrane protease

serine 6), have been associated with low iron status [18–20] and a condition described as iron-refractory iron deficiency anaemia (IRIDA) [18–21]. Individuals with IRIDA have a hereditary form of anaemia that does not respond to oral iron supplementation [22,23]. Although IRIDA is quite rare, it may be at the extreme end of a broad continuum of disease, since *TMPRSS6* genetic variants can lead to different degrees of iron deficiency and anaemia [18–20]. In addition, SNPs in the *TF* gene, also important in iron transport to cells, have also been reported to affect iron status and lead to low iron status [24–26]. Furthermore, SNPs in the divalent metal transporter 1 (*DMT1*), the duodenal apical iron transporter encoded by the *SLC11A2* gene have been associated with an unusual syndrome characterized by microcytic anaemia and a paradoxical iron overload [27,28].

A genome-wide association study (GWAS) investigating genetic determinants of relevant haematological traits and iron status have identified variants in *TF* and *HFE,* which explain approximately 40% of variation in serum transferrin levels [26]. Also, GWASs have identified genetic variants in *TMPRSS6* associated with alterations of serum iron status, erythrocyte volume [29], and haemoglobin levels [20]. African populations have been greatly under-represented in such studies. A GWAS using an African population cohort replicated only the association of two SNPs in *TMPRSS6* with lowered haemoglobin concentration, and one SNP in *TF* with increased ferritin concentrations [30]. Differences in the frequencies of risk alleles and linkage disequilibrium patterns might explain the limited replication of association results between European, Asian and African populations. Hence, there is a need to investigate population-specific genetic variants that may affect iron status.

Here, we conducted a review of the literature to identify genetic variants that have been associated with iron imbalances, with a special focus on SNPs in *TMPRSS6, HAMP, TF, TFR2, SLC40A1* and *HFE* genes. We investigated the geographical distribution of studies and assessed the differences in allele frequency of these polymorphisms and their linkage disequilibrium patterns across global populations. We use genetic data from our Keneba Biobank in rural Gambia and from the 1000 Genomes Project. We also explored the possibility of natural selection acting on these genes and any resulting population-specific selection, as measured through large differences in allele frequencies between geographic regions. As part of this, we sought to summarize the geographical distribution of genetic determinants of iron status. The resulting insights may assist in designing future genetic association studies that are geared towards identifying population-specific genetic risk factors affecting iron status and, ultimately, guiding population-specific iron intervention strategies.

Methods

Selection of SNPs

A literature search was conducted using the Human Genetic Epidemiology (HuGE) navigator, a database of published population-based human genetic epidemiology studies. This review was complemented using the PubMed site with search terms: “anaemia”, “iron”, “iron overload”, “iron deficiency anaemia”, “iron imbalance”, “hepcidin”, “genome-wide association study”, “GWAS”, “haematology traits”, and “haemochromatosis”. The search was conducted on articles published between 01 January 1999 to 31 October 2018. The assessment process included examining titles and abstracts of studies and excluding duplicates. Articles were included if they were:

(1) original research papers conducted in humans; (2) tested for an association between at least one SNP in the genes commonly linked to dysregulated iron status (*TMPRSS6, HAMP, TF, TFR2, SLC40A1* and *HFE*) or iron status measures. These include iron status biomarkers (serum iron, transferrin, ferritin, soluble transferrin receptor, transferrin saturation, total iron binding capacity, unsaturated iron binding capacity and hepcidin) alone or in combination with haematology traits (haemoglobin, red blood cells, hematocrit, mean corpuscular haemoglobin and mean corpuscular hemoglobin concentration). Animal studies, case reports, commentaries and articles not written in English were excluded. Rare variants reported in a single individual or family were discarded. Information on genomic and gene location, allele ancestry, minor allele variant and the predicted consequence of each SNP were obtained from the Ensembl dataset (release 98) [31] and the dbSNP nucleotide variation database [32].

Genotype data and statistical analysis

We obtained genotype data from the Keneba at MRCG at LSHTM [33] (n=3,116 healthy Gambian individuals) and from the 1000 Genomes project [n=2,504; 26 populations categorised into African (AFR, n=661), European (EUR, n=503), American (AMR, n=347), East Asian (EAS, n=504) and South Asian (SAS, n=489)] [34]. Genotyping of the Keneba Biobank populations was performed using the Infinium 240K Human Exome Beadchip (v1.0 and v1.1). Genotype calling was performed using data-driven clustering (Genome Studio, Illumina, CA, USA).

We assessed the differences in allele frequencies for SNPs with genotype calls in both the Keneba Gambian and the pan-African populations in the 1000 Genomes Project.

Linkage disequilibrium (LD) measures (*D'* and r^2) were calculated using the R package Genetics [35]. The correlation between minor allele frequencies across populations was calculated using the Pearson's correlation coefficient in the R package corrplot. We calculated the allele risk score for each individual by aggregating the number of risk alleles an individual carried. To do this, from each SNP, the risk allele was assigned 1 and alternate allele assigned 0. For the genotype of each SNP, an individual was given either 0 (wildtype), 1 (heterozygote) or 2 (homozygote for the risk allele). Using this information, we determined the allele risk scores across populations for both low and high iron SNPs. For 23 SNPs it was not possible to identify the associated alleles (e.g. just a "A/T" label) or classify the direction of association (e.g. absence of regression coefficients). Also, for some SNPs (*TF* rs3811658 and rs1880669, and *TMPRSS6* rs2072860 and rs2111833) (**S1 Table**) we found contradictory information about their association with iron biomarkers between studies. They were all excluded from risk allele analysis. Statistic differences in the distribution of risk alleles between populations were calculated using a Wilcoxon rank sum test in the R statistical package [36]. To allow for multiple comparisons, a Bonferroni correction was applied.

The minor allele frequency (MAF), observed and expected heterozygosities and measures of population differentiation (global and pairwise F_{ST} to assess differences in allele frequencies) were calculated from the genotype data for all iron-associated SNPs using a combination of the R packages Adegenet [37], Hierfstat [38] and Pegas [39]. Weir & Cockerham F_{ST} values were calculated and range from 0 to 1, where a zero value implies that the two populations are interbreeding, and a value of one means that the two populations do not share any genetic diversity. Population Branch

Statistic (PBS) values were calculated using the F_{ST} data from the comparison of three populations (AFR-EUR, AFR-SAS, EUR-SAS) according to methods described elsewhere [40]. To evaluate the significance of the observed F_{ST} and PBS values, the results were compared with the empirical distribution of genome-wide SNPs reported by others using individuals from several geographical locations and including data from the HapMap and HGDP [41–45].

Statistical differences between MAFs were analysed using the two-proportion Z-Test in R. The integrated Haplotype Score (iHS) [46,47] statistic was investigated using Haplotter (http://haplotter.uchicago.edu/)[48] and HGDP selection browsers (http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/) at the individual genes and surrounding regions.

Ethics Statement

The Keneba Biobank Project received ethical approval from the MRCG at LSHTM Scientific Coordinating Committee and the MRCG at LSHTM/ Gambia Government Joints Ethics Committee (SCC1185). Written informed consent was obtained from each participant.

Results

Genetic variants associated with iron imbalances

A total of 64 studies were selected that contained data on the effects of genetic polymorphisms on the variations in iron or haematological parameters (**S1 Table).** The majority of the studies (59/64) were conducted in Europe, Asia and the USA (**Fig 1, S2 Table**). Only five studies were conducted in Africa, two in Rwanda [49,50], one in

Zimbabwe [51], one in South Africa [52] and one meta-analysis across Kenya, Tanzania and South Africa [30]. Across the 64 studies, 50 SNPs were identified in six genes (*TMPRSS6, HAMP, TF, TFR2, SLC40A1* and *HFE)* (**S1 Table)**. More than half of these SNPs were found to be associated with variation in iron or in other haematological parameters in more than one country (29 SNPs, 58%). Of these 29 SNPs, 79.3% were reported in more than one ethnic group (**S2 Table**). Nine SNPs lead to a missense mutation causing an amino acid change, four SNPs had synonymous variants, and the remaining SNPs are in intronic (n=32), regulatory or intergenic regions (n=5).

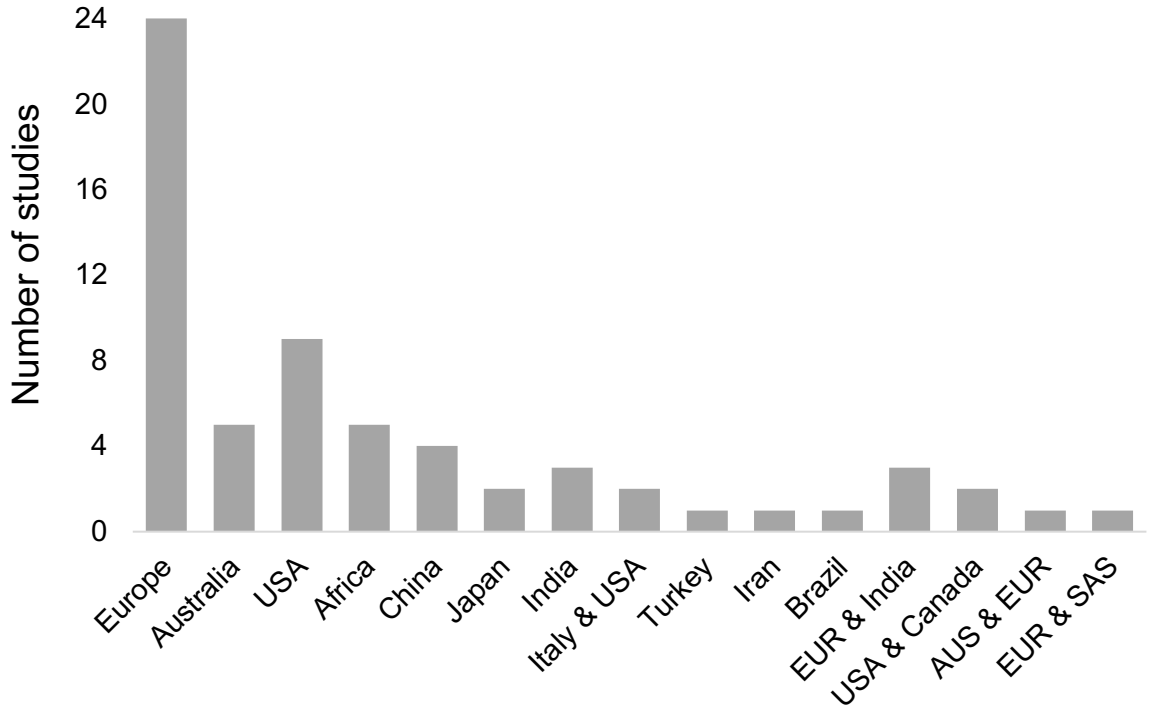


Figure 1. Geographical locations of the sixty-four studies that reported genetic variants associated with iron imbalance. Nine studies involved multi-ethnic populations. AUS, Australia; EUR, Europe; SAS, South Asia.

The highest number of SNPs were identified in the *TMPRSS6* gene region (n=23), where the majority were associated with IRIDA, iron deficiency or indicators of low iron status (**S1 Table**). The most commonly reported *TMPRSS6* SNP was rs855791, followed by rs4820268, rs2235321 and rs2235324, all associated with biomarkers of low iron status. These SNPs have been mainly reported in non-African populations. Three *TMPRSS6* SNPs (rs5756504, rs5756506 and rs1421312) were also associated with biomarkers indicating elevated iron status (**S1 Table**).

The *TF* gene had the second highest number of SNPs related to either low or high iron status (n=18). The most common of these (rs3811647) was reported by ten studies (**S1 Table**). This variant has been mainly associated with elevated transferrin and total iron binding capacity levels [26,53,54]. For the *SLC40A1* gene, three SNPs were selected that led to alterations in iron status measures and severity of haemochromatosis [50,51,55–57]. One SNP was identified in *HAMP* (rs10421768) [30,55,58–60] and one in *TFR2* (rs7385804) (6,7,33,34,37,38), both of which were found to be associated with increases in haemoglobin and alterations serum in ferritin concentrations [30,55,56,58,61–63]. For the *HFE* gene, we found four SNPs that have been associated with alterations in haemoglobin and/or an increase in the genetic risk of hereditary haemochromatosis [13,14,19,20,24,26,29,56,62,64–74]. The most commonly reported *HFE* variant is rs1800562 (C282Y) [13,19,24,26,29,71,72], which has been widely associated with the severe form of hereditary haemochromatosis in European descents.

Global geographic distribution of allele frequencies

We investigated the allele frequencies of the 50 SNPs across data from the Keneba Biobank at the MRCG at LSHTM in The Gambia (n= 3,116) and the 1000 Genome

project (n= 2,504) [34]. The 1000 Genomes project includes data from African (AFR, n=661; including from The Gambia), European (EUR, n=503), American (AMR, n=347), East Asian (EAS, n=504) and South Asian (SAS, n=487) populations. Only thirteen of the 50 SNPs in the *TF*, *TMPRSS6*, *HFE* and *SLC40A1* genes, were available in the Keneba Biobank, because not all the SNPs were on the Exome chip that was used for genotyping this population. When we compared the allele frequencies of the SNPs with data from The Gambians in the Keneba Biobank with the pan-African populations in the 1000 Genomes project, we observed minimal differences (**Fig 2**).

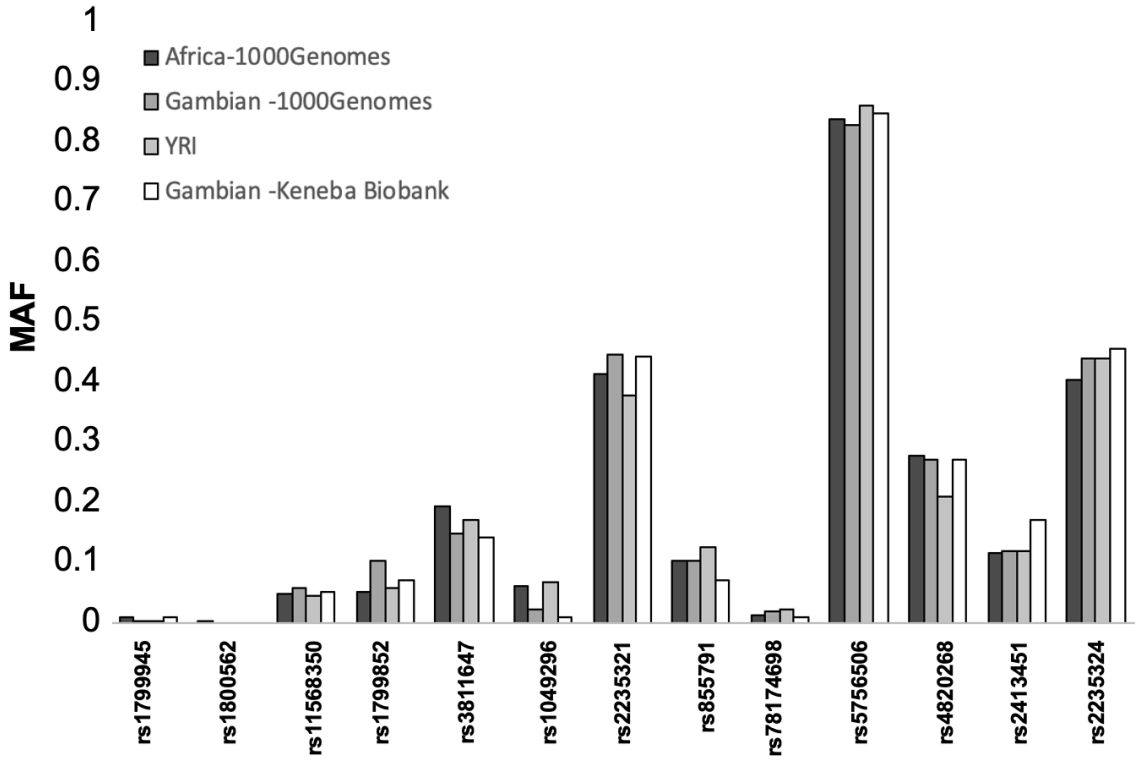


Figure 2. Minor Allele frequencies (MAF) of 13 SNPs across African populations. Comparing MAF between the two Gambian datasets, Yoruba (YRI) from Nigeria and overall African populations included in the 1000 Genomes Project. The minor alleles were defined by the 1000 Genomes Project.

For the majority of SNPs, the MAFs in the African populations were very different to other worldwide populations (**Fig 3** and **Fig 4)**. The greatest allele frequency differences were observed in rs1439816 in *SLC40A1*, and in several SNPs in *TMPRSS6* (including rs855791 and rs855788). The intronic variant rs1439816 in the *SLC40A1* gene has a MAF of ~20% in the non-African populations but reaches >73% frequency in Africa (**S1 Table**). The missense variant A736V (*TMPRSS6* rs855791) is the most reported SNP associated with iron deficiency and has a MAF of ~50% across all non-African populations, but in Africa it only reaches 10% (7% in the MRCG Keneba Biobank population) (**Fig 4**). The intronic variant rs855788 in *TMPRSS6* has a MAF of ~30% across non-African populations, contrasting with a frequency in excess of 86% in the African populations (**Fig 4**).

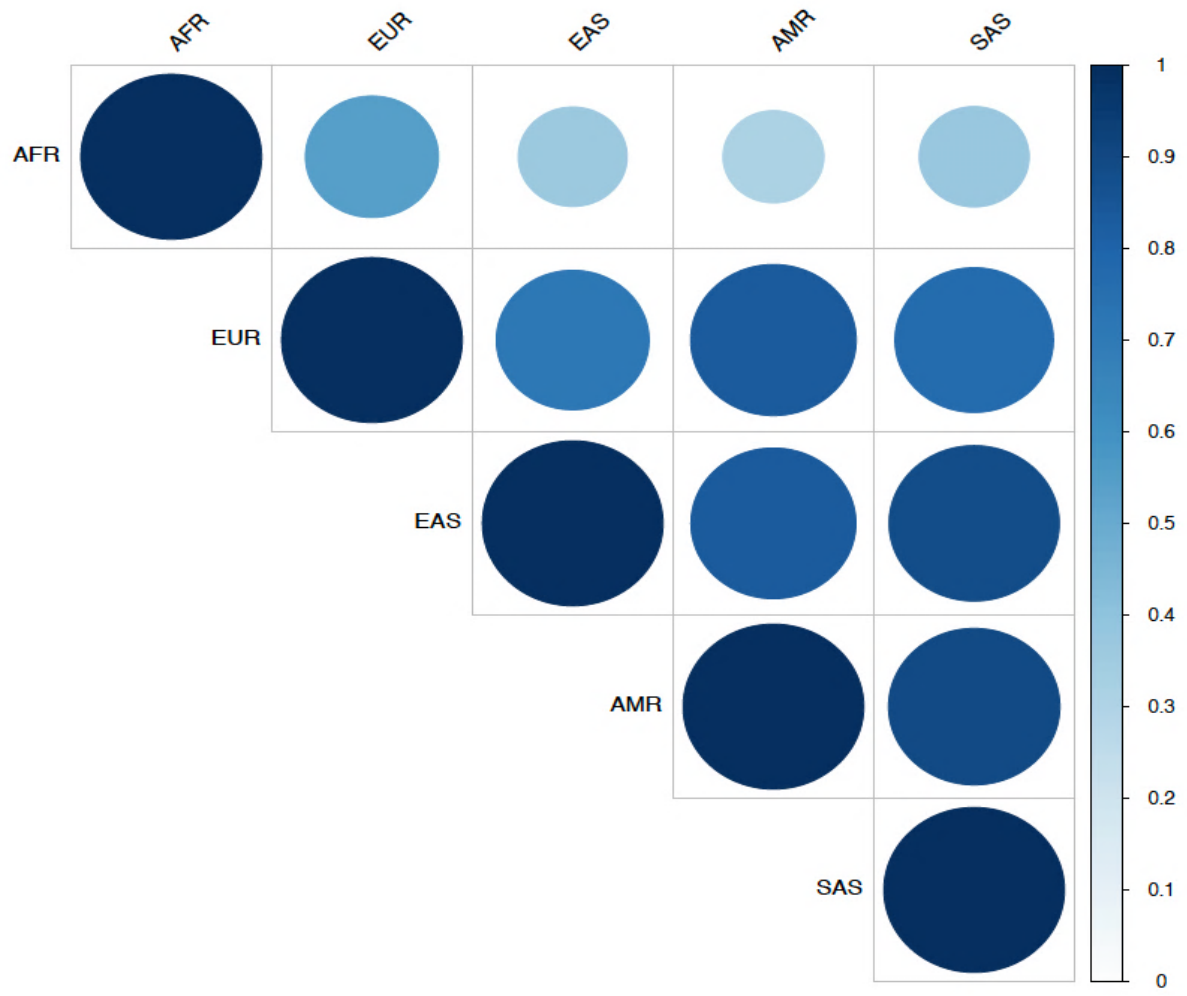
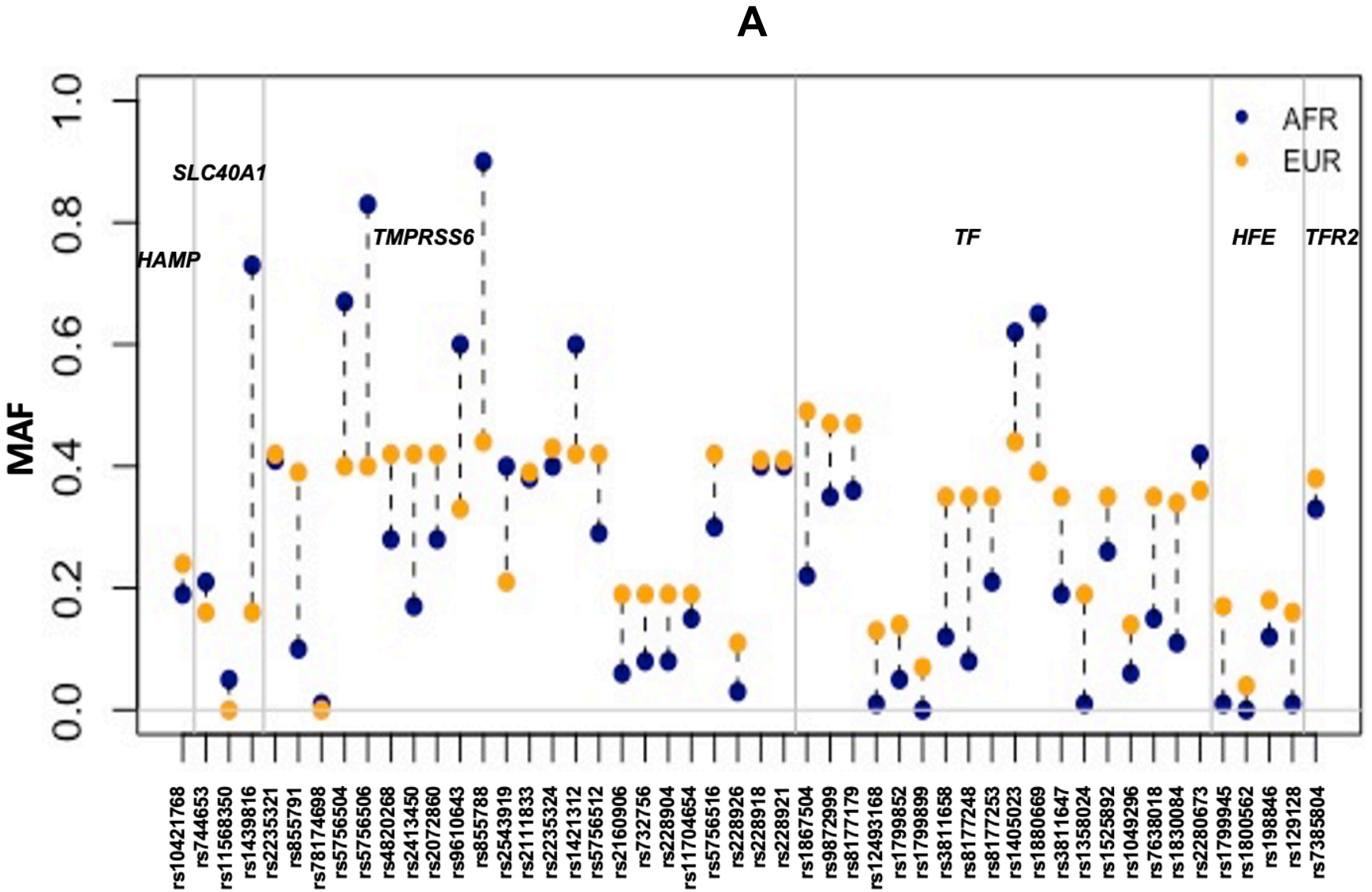
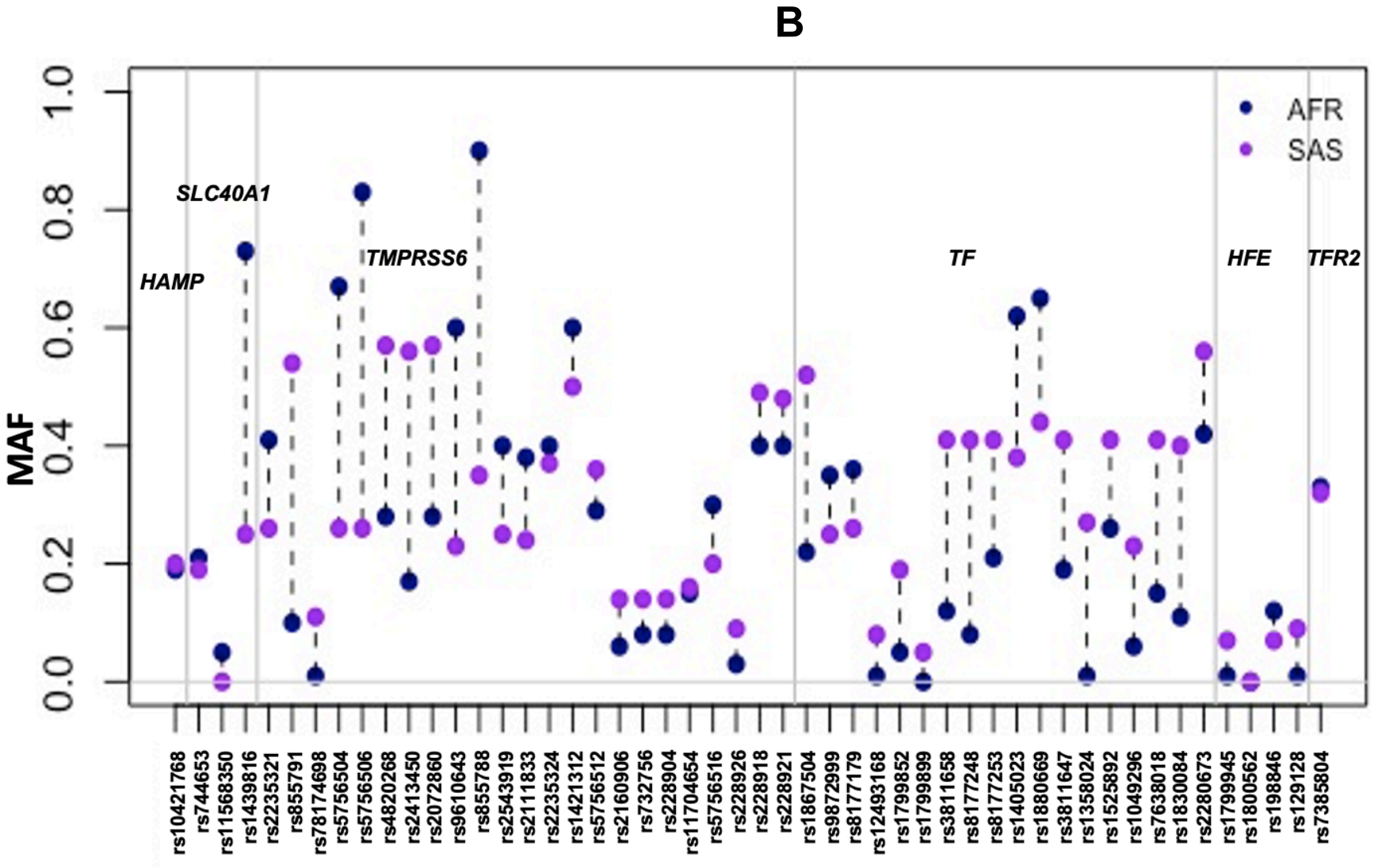
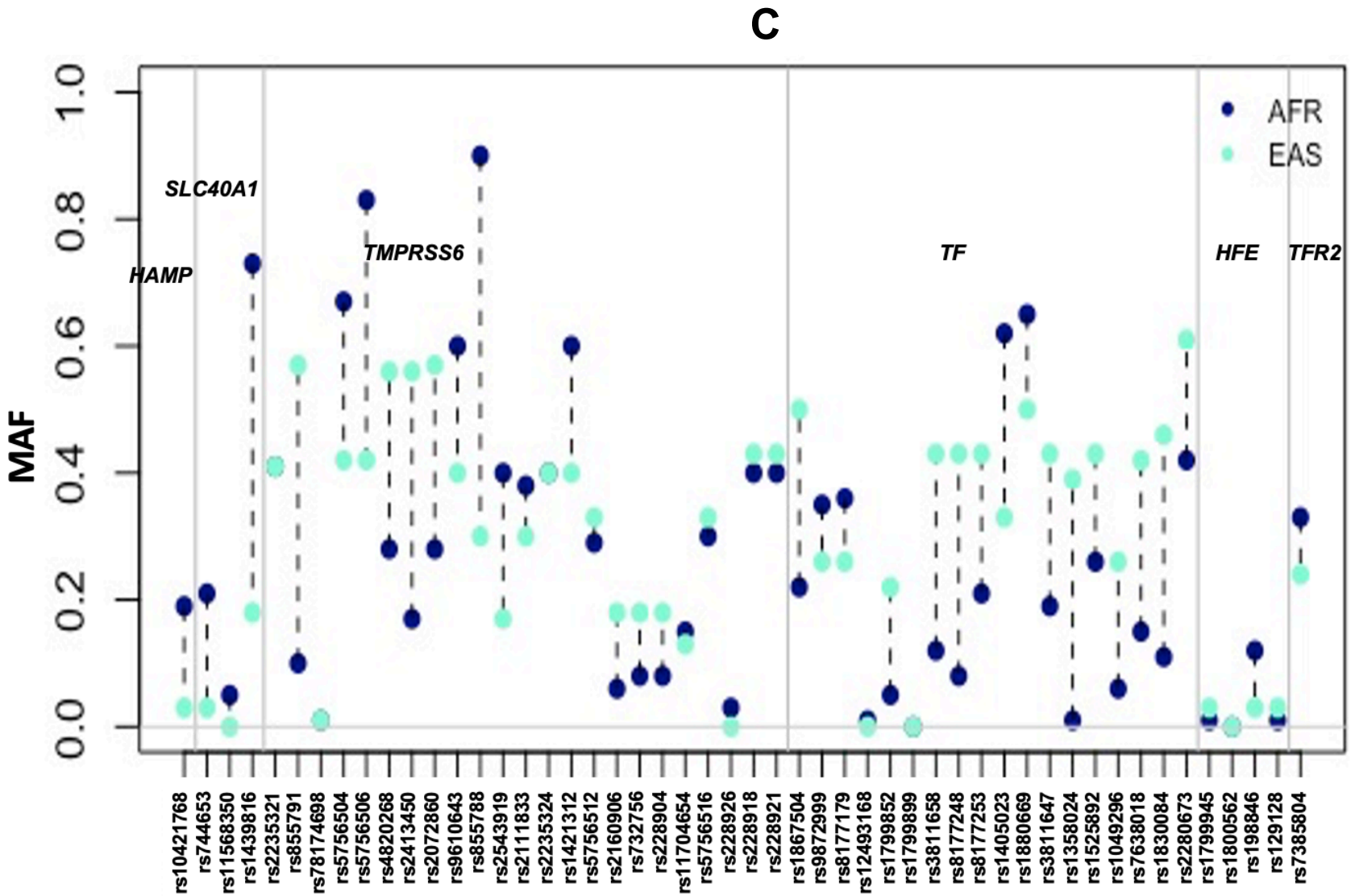


Figure 3. Correlation of minor allele frequencies between different geographic regions. Correlation coefficients were obtained by pairwise comparisons of each of the 50 SNPs identified across two populations. They are coloured according to the value using a gradient from white (representing 0 for no correlation) to dark blue (1 for perfect correlation). The minor allele variant was defined by the 1000 Genomes Project. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.


A
MAF
0.0
0.2
0.4
0.6
0.8
1.0
AFR
EUR
SLC40A1
HAMP
TMPRSS6
TF
HFE
TFR2
rs10421768
rs744653
rs11568350
rs1439816
rs2235321
rs855791
rs78174698
rs5756504
rs5756506
rs4820268
rs2413450
rs2072860
rs9610643
rs855788
rs2543919
rs2111833
rs2235324
rs1421312
rs5756512
rs2160906
rs732756
rs228904
rs11704654
rs5756516
rs228926
rs228918
rs228921
rs1867504
rs9872999
rs8177179
rs12493168
rs1799852
rs1799899
rs3811658
rs8177248
rs8177253
rs1405023
rs1880669
rs3811647
rs1358024
rs1525892
rs1049296
rs7638018
rs1830084
rs2280673
rs1799945
rs1800562
rs198846
rs129128
rs7385804


B
MAF
0.0
0.2
0.4
0.6
0.8
1.0
AFR
SAS
SLC40A1
HAMP
TMPRSS6
TF
HFE
TFR2
rs10421768
rs744653
rs11568350
rs1439816
rs2235321
rs855791
rs78174698
rs5756504
rs5756506
rs4820268
rs2413450
rs2072860
rs9610643
rs855788
rs2543919
rs2111833
rs2235324
rs1421312
rs5756512
rs2160906
rs732756
rs228904
rs11704654
rs5756516
rs228926
rs228918
rs228921
rs1867504
rs9872999
rs8177179
rs12493168
rs1799852
rs1799899
rs3811658
rs8177248
rs8177253
rs1405023
rs1880669
rs3811647
rs1358024
rs1525892
rs1049296
rs7638018
rs1830084
rs2280673
rs1799945
rs1800562
rs198846
rs129128
rs7385804


C
MAF
0.0
0.2
0.4
0.6
0.8
1.0
AFR
EAS
SLC40A1
HAMP
TMPRSS6
TF
HFE
TFR2
rs10421768
rs744653
rs11568350
rs1439816
rs2235321
rs855791
rs78174698
rs5756504
rs5756506
rs4820268
rs2413450
rs2072860
rs9610643
rs855788
rs2543919
rs2111833
rs2235324
rs1421312
rs5756512
rs2160906
rs732756
rs228904
rs11704654
rs5756516
rs228926
rs228918
rs228921
rs1867504
rs9872999
rs8177179
rs12493168
rs1799852
rs1799899
rs3811658
rs8177248
rs8177253
rs1405023
rs1880669
rs3811647
rs1358024
rs1525892
rs1049296
rs7638018
rs1830084
rs2280673
rs1799945
rs1800562
rs198846
rs129128
rs7385804

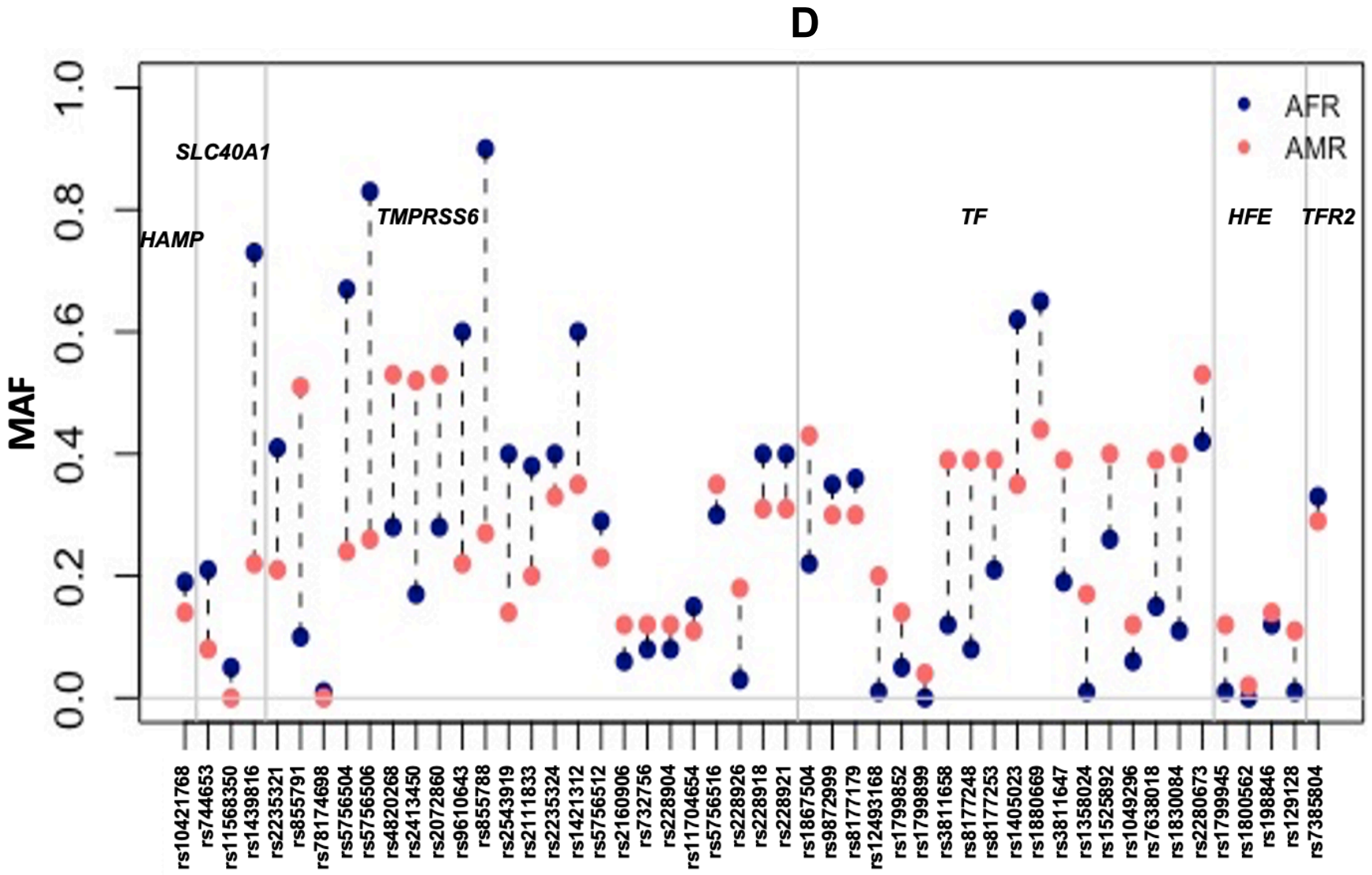
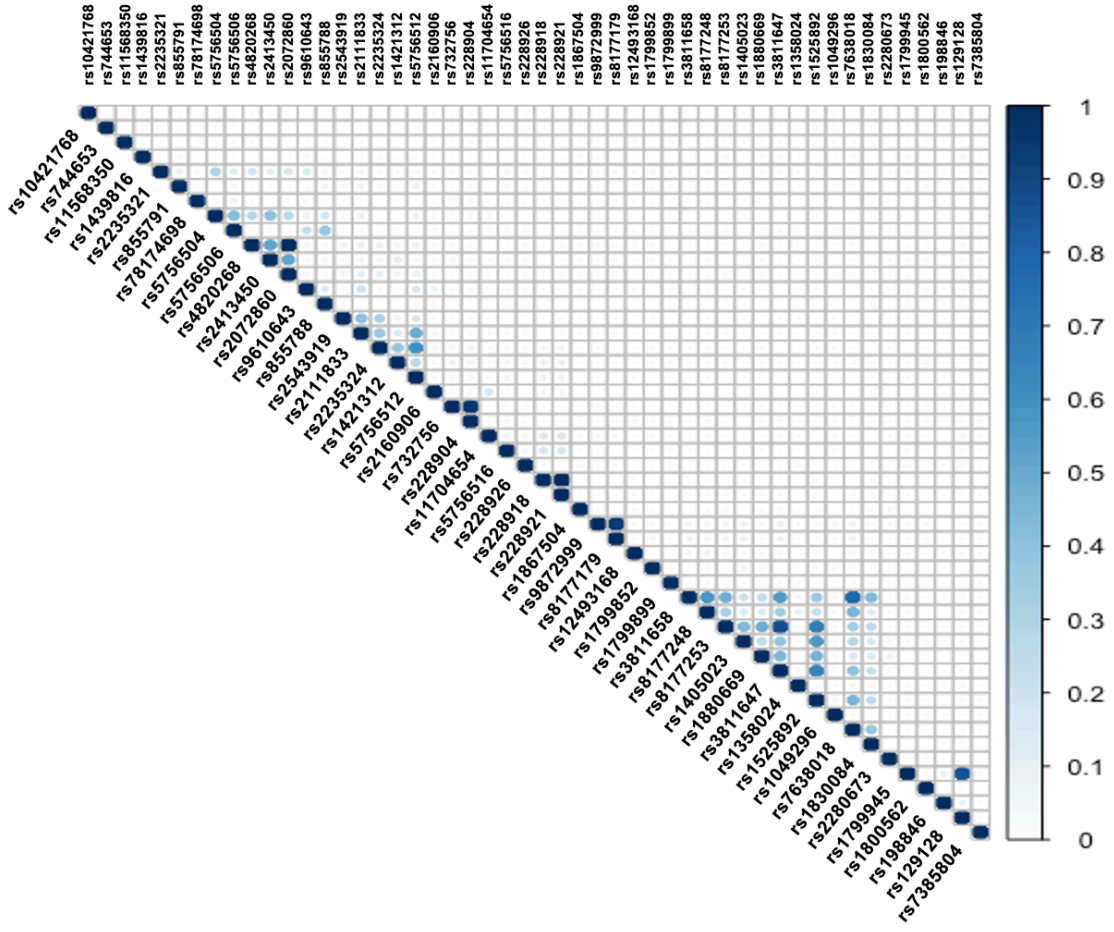

D
MAF
0.0
0.2
0.4
0.6
0.8
1.0
AFR
AMR
SLC40A1
HAMP
TMPRSS6
TF
HFE
TFR2
rs10421768
rs744653
rs11568350
rs1439816
rs2235321
rs855791
rs78174698
rs5756504
rs5756506
rs4820268
rs2413450
rs2072860
rs9610643
rs855788
rs2543919
rs2111833
rs2235324
rs1421312
rs5756512
rs2160906
rs732756
rs228904
rs11704654
rs5756516
rs228926
rs228918
rs228921
rs1867504
rs9872999
rs8177179
rs12493168
rs1799852
rs1799899
rs3811658
rs8177248
rs8177253
rs1405023
rs1880669
rs3811647
rs1358024
rs1525892
rs1049296
rs7638018
rs1830084
rs2280673
rs1799945
rs1800562
rs198846
rs129128
rs7385804

Figure 4. The differences in minor allele frequencies of SNPs in the six genes investigated, across different geographic regions. The comparisons were made between Africans and other global populations (A) Africa vs. Europe; (B) Africa vs. South Asia; (C) Africa vs. East Asia; (D) Africa vs. America. The thick grey lines indicated borders between SNPs in different genes: *HAMP*, *SLC40A1, TMPRSS6, TF, HFE* and *TFR2.* The minor alleles were defined according to the 1000 Genomes Project database [34]. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.

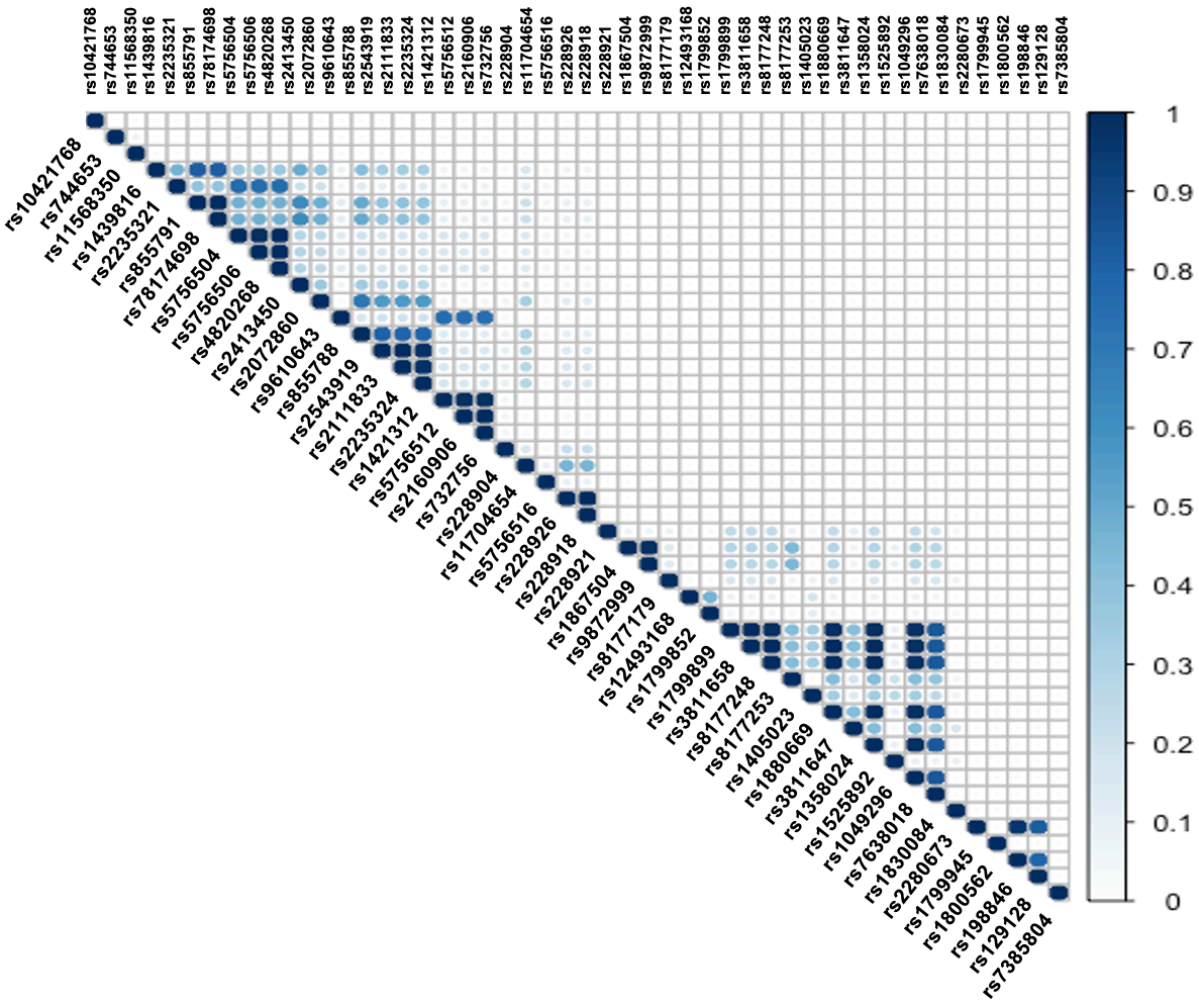
From the selected SNPs, several in African (n=10 SNPs) and East Asian (n=11 SNPs) populations have fixed ancestral alleles or low MAF (<5%) (**Fig 4**). These SNPs include four missense variants, with the lowest overall MAF or with fixed ancestral alleles in several populations (associated with low iron: *TMPRSS6* rs78174698 and *TF* rs1799899; associated with increased serum ferritin: SLC40A1 rs11568350; associated with haemochromatosis: *HFE* rs1800562). The *TMPRSS6* rs78174698 (P555S) MAF is low overall (<2%) across most populations, except in South Asia where the minor allele is >10%. The minor allele for rs1799899 (G277S) is rare in Africa and East Asia (<0.2%), and only reaches >4% MAF in European, American and South Asian populations. For *SLC40A1* rs11568350 (Q248H), the minor allele reaches 5% in Africans, including in both The Gambian populations in the two datasets analysed. In the other global populations, the ancestral allele is almost fixed. The variant A allele of rs1800562 (C282Y) has the highest frequency in European populations (4.3% and 5.3% in Caucasians from Europe in the 1000 Genome Project and in the HapMap CEU population, which have ancestry from Northern and Western Europe, respectively). The frequency of this variant is extremely low in Africans (0.2% in the 1000 Genomes project) and it was not detected (MAF=0) in the Keneba Biobank population.

We also investigated the population-specific linkage disequilibrium (LD) patterns between SNPs in the candidate genes. There were blocks of high LD in the non-African population, and the overall levels of LD were lower in the African populations **(Fig 5, Fig S1)**, including in The Gambia. In contrast, the SNPs in the *TF* gene still showed a pattern of high LD in the African populations.

A


rs10421768
rs744653
rs11568350
rs1439816
rs2235321
rs855791
rs78174698
rs5756504
rs5756506
rs4820268
rs2413450
rs2072860
rs9610643
rs855788
rs2543919
rs2111833
rs2235324
rs1421312
rs5756512
rs2160906
rs732756
rs228904
rs11704654
rs5756516
rs228926
rs228918
rs228921
rs1867504
rs9872999
rs8177179
rs12493168
rs1799852
rs1799899
rs3811658
rs8177248
rs8177253
rs1405023
rs1880669
rs3811647
rs1358024
rs1525892
rs1049296
rs7638018
rs1830084
rs2280673
rs1799945
rs1800562
rs198846
rs129128
rs7385804
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

B


rs10421768
rs744653
rs11568350
rs1439816
rs2235321
rs855791
rs78174698
rs5756504
rs5756506
rs4820268
rs2413450
rs2072860
rs9610643
rs855788
rs2543919
rs2111833
rs2235324
rs1421312
rs5756512
rs2160906
rs732756
rs228904
rs11704654
rs5756516
rs228926
rs228918
rs228921
rs1867504
rs9872999
rs8177179
rs12493168
rs1799852
rs1799899
rs3811658
rs8177248
rs8177253
rs1405023
rs1880669
rs3811647
rs1358024
rs1525892
rs1049296
rs7638018
rs1830084
rs2280673
rs1799945
rs1800562
rs198846
rs129128
rs7385804
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

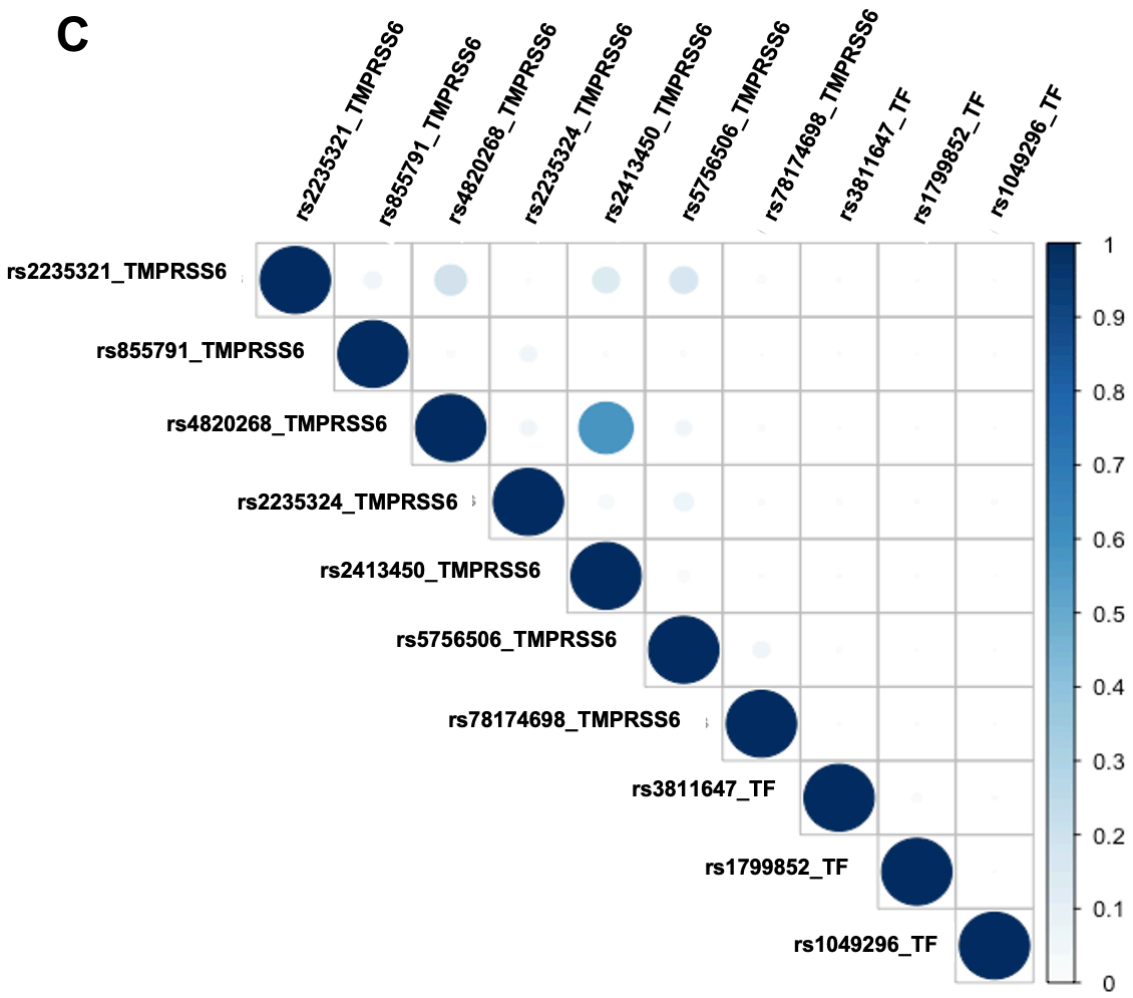


Figure 5. Linkage disequilibrium (LD) plots in SNPs in *HAMP, SLC40A1*, *TMPRSS6*, *TF*, *HFE* and *TFR2* genes. LD plot showing r^2 values in SNPS associated with iron imbalances in: (A) African populations, (B) European populations and (C) Gambian population in the Keneba Biobank.

Distribution and frequency of iron imbalance risk alleles

To investigate if any population had an over- or under-representation of risk alleles leading to iron imbalances, we first classified the alleles as protective or susceptible based on previous associations with low or high iron status or related biomarkers (**S1 Table**). A total of 23 SNPs were included in the risk allele analysis (see Methods for exclusion criteria). Eleven SNPs had alleles that were clearly associated with low iron, iron deficiency anaemia and/or IRIDA (SNPs in *TMPRSS6* (rs855791, rs2235321,

rs2235324, rs4820268, rs2413450, rs228916, rs228918 and rs228921) and *TF* (rs3811647, rs1799899 and rs8177253) (**S1 Table**).

The South and East Asian populations had the highest number of low iron risk alleles, whereas, Africans had the lowest and were significantly different from the other populations (**Fig 6A**, P < 0.0001). The American and European populations had similar number of low iron risk alleles, but lower than the Asian populations (P <0.0001) (**Fig 6A**).

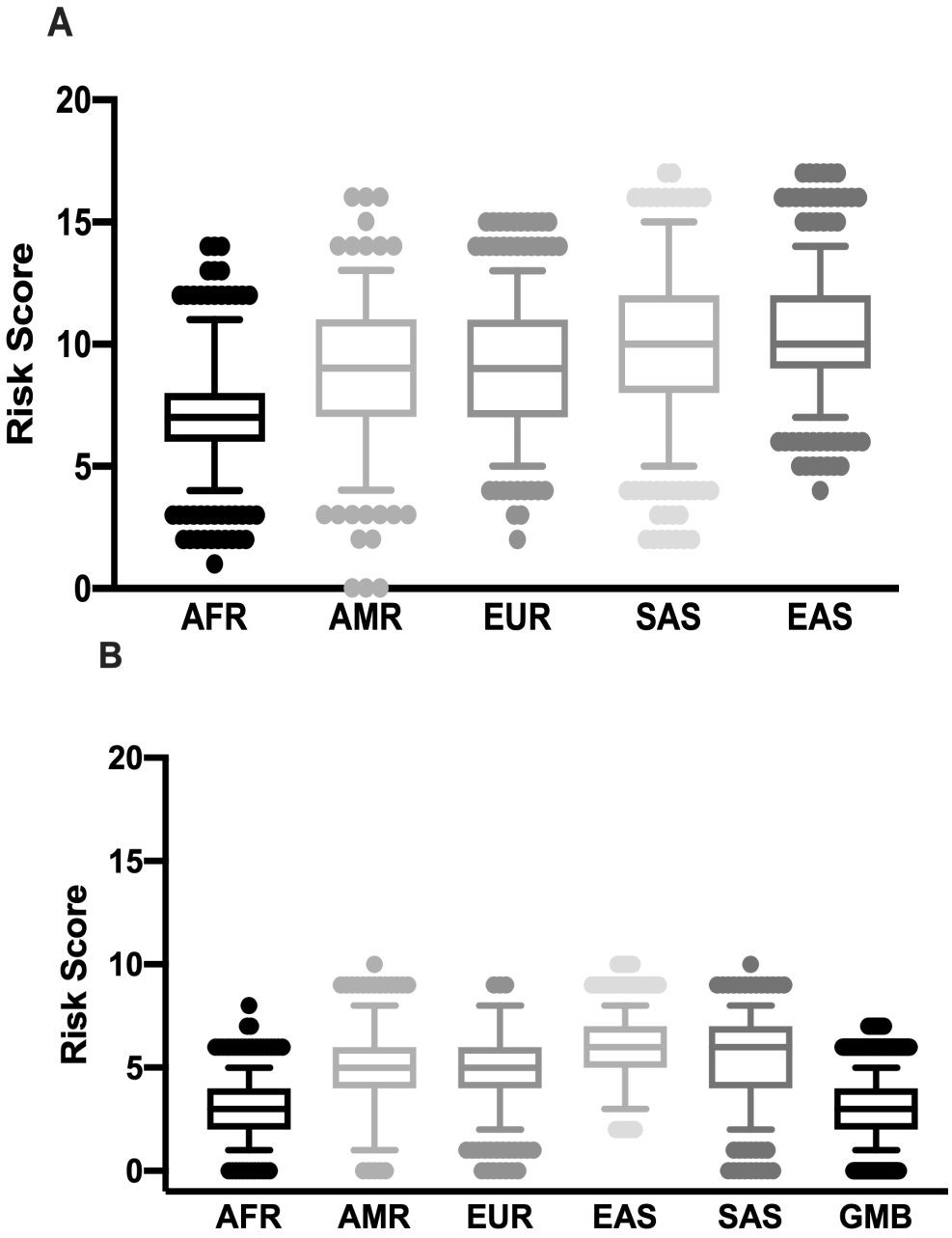


Figure 6. Distribution of the number of low iron risk alleles across global populations. (A) Distribution of the number of low iron risk alleles in eleven SNPs associated with low iron status across five populations. (B) Distribution of the number of low iron risk alleles in six SNPs with genotype data in the MRC Keneba Biobank population. Designation of the allele (risk or not) was determined by their previously published information as presented in S1 and S2 Tables. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.

Out of the eleven SNPs we found to be associated with low iron, it was only possible to compare six using the Keneba Biobank data, as data on the remaining SNPs were not available. The number of low iron risk alleles of the Gambians in the Keneba Biobank and the overall Africans in the 1000 Genomes were similar (**Fig 6B**). However, the low iron risk alleles in the Gambian and overall African populations were significantly lower compared to the other populations ($P < 2x10^{-16}$) (**Fig 6B)**.

Twelve SNPs were clearly associated with high iron or related biomarker (SNPs in *HAMP, TMPRSS6, TF, SLC40A1, TRF2* and close to *HFE*) with their risk alleles indicated (**S1 Table**). Three out of these twelve high iron associated SNPs were in or close to the *HFE* gene (rs1799945, rs1800562 and rs198846). These three SNPs were associated with haemochromatosis. Since haemochromatosis is predominantly common in those of European descent and rare in other populations, we analysed these SNPs separately. The European populations have the highest number of high iron risk alleles, significantly different from the other populations ($P < 0.00850$) (**Fig 7)**.

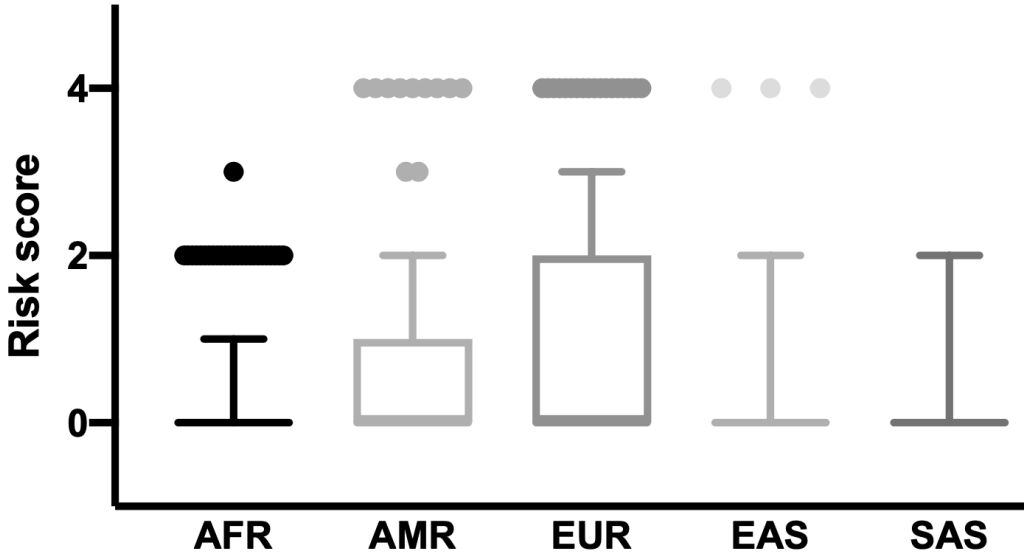


Figure 7. Distribution of the number of risk alleles for haemochromatosis among global populations. Designation of the risk allele was determined by previously published information as presented in S1 and S2 Tables. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.

Data for two of the SNPs (rs1799945 and rs1800562) were available for the Keneba Biobank population, but the frequency of risk alleles was low (1% and 0%, respectively). Therefore, we could not compare the frequencies of risk alleles of these SNPs between the Keneba Biobank population and the 1000 Genomes project populations. Furthermore, we compared the frequencies of the high iron risk alleles of the remaining nine SNPs associated with elevated iron status in other genes. The African population in the 1000 Genomes Project had a significantly lower number of high iron risk alleles than the other populations ($P < 0.0001$) (**Fig 8A**). The distributions between the other populations were similar. From these nine SNPs, genotype data for three SNPs (*TMPRSS6* rs5756506, *TF* rs1799852 and *SLC40A1* rs11568350 (Q248H) were available for the Gambians in the Keneba Biobank. When we compare the frequencies of the high iron risk alleles at these three SNPs across populations

(**Fig 8A**), Gambians in the Keneba Biobank and pan-African populations have the lowest number of combined risk alleles for high iron (**Fig 8B**).

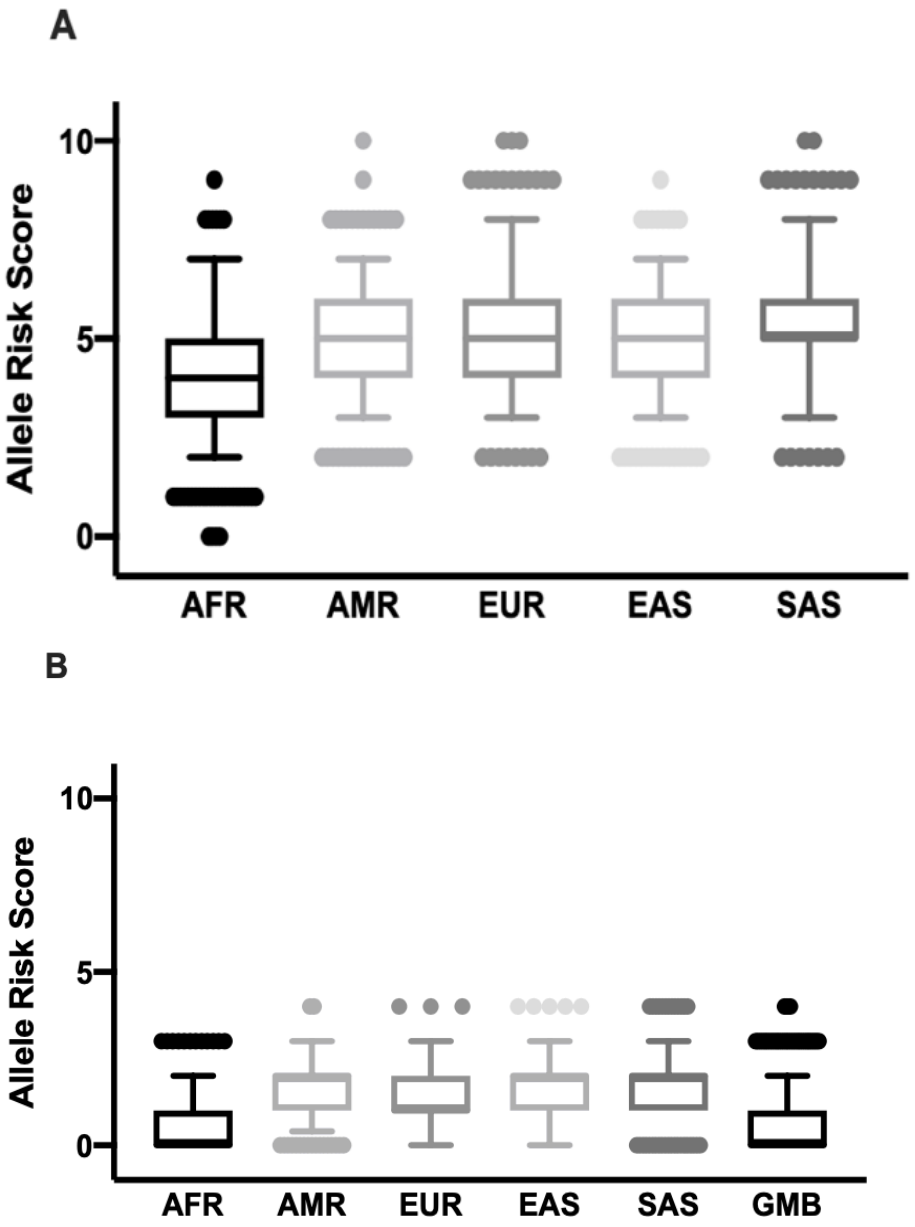


Figure 8. Distribution of the number of high iron risk alleles across global populations. (A) Distribution of the number of high iron risk alleles across five populations. (B) Distribution of the number of high iron risk alleles in three SNPs in the MRC Keneba Biobank (Gambian) and other populations. Designation of the allele (risk or not) was determined by their previously published information as presented in S1 and S2 Tables. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.

Global population differentiation

We calculated the global and pairwise fixation index (F_{ST}) across the 5 populations to assess population divergence for all iron-associated SNPs. The overall F_{ST} across the populations was 0.076. The pairwise F_{ST} between the continental groups shows that African versus non-African populations had the greatest allele frequency differentiation (F_{ST} >0.09; **Table 1**).

Table 1. Pairwise F_{ST} values between populations

	EUR	EAS	AMR	SAS
EAS	0.0317			
AMR	0.0248	0.0232		
SAS	0.0263	0.0154	0.0130	
AFR	0.0992	0.1465	0.1507	0.1425

AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian; $F_{ST,}$ fixation index.

We then investigated the individual SNPs driving the differentiation between African and other populations (**Fig 9**). The variants with the highest F_{ST} (>0.3) and highest allele frequency differences were rs1439816 in *SLC40A1* and rs855791, rs855788 and rs5756506 in *TMPRSS6* (**Fig 9).** The average F_{ST} values for the set of SNPs in each population was less than 0.065. The highest F_{ST} values we observed lay within the top 5% of the distribution of empirical global F_{ST} values described by others (95% percentile F_{ST} > 0.28) [43–45].

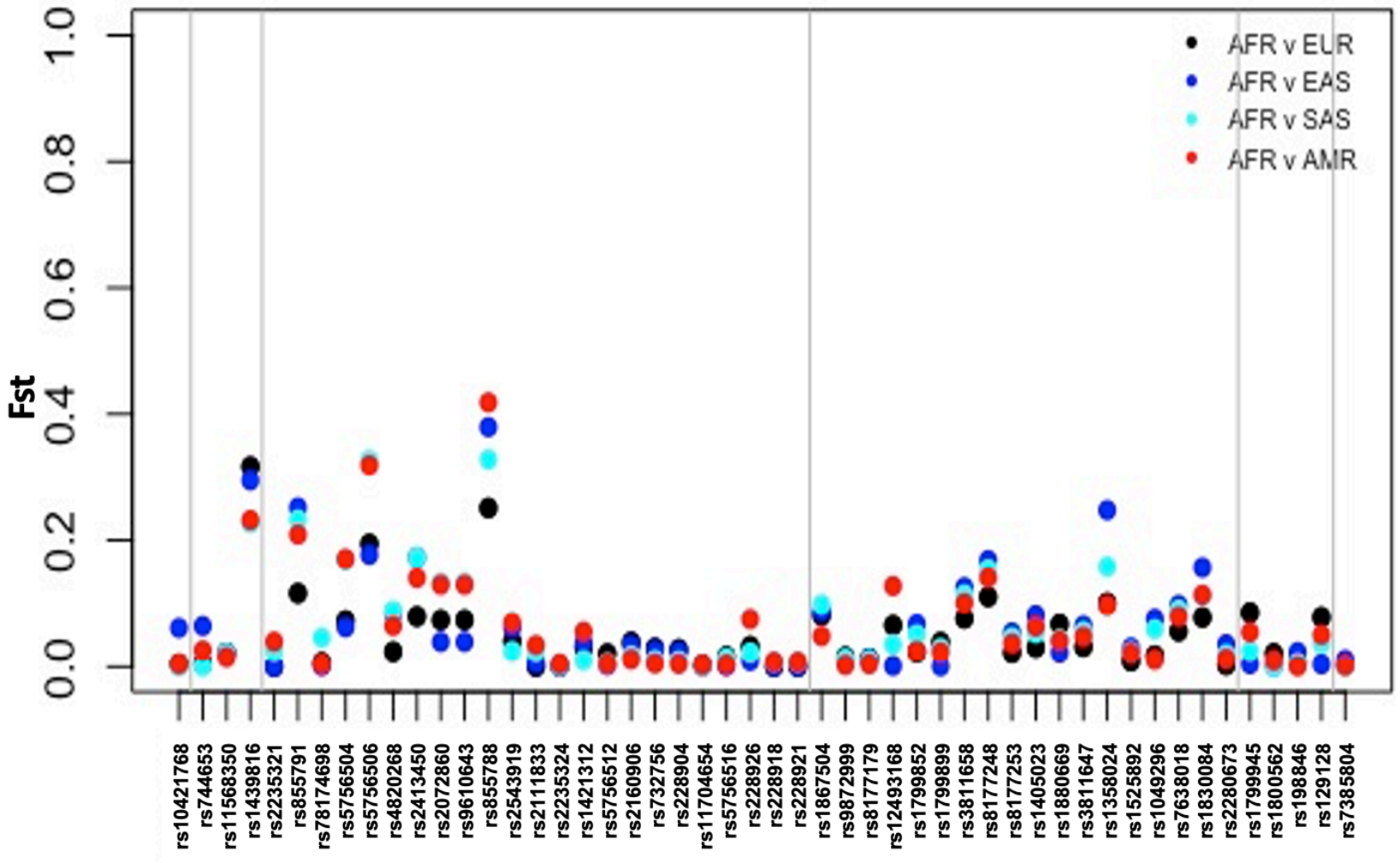


Figure 9. Pairwise F_{ST} values for iron related SNPs between African and non-African populations. This figure illustrates the comparison of F_{ST} scores between African and other global populations. AFR, African; AMR, American; EUR, European; EAS, East Asian; SAS, South Asian, F_{ST}, fixation index.

We also calculated the Population Branch Statistic (PBS) values, an F_{ST}-based test involving the comparison of three populations, to investigate if the differentiation between populations could be driven by positive selection [40]. We used African, European and South Asian populations and observed that the PBS analysis reaffirms the F_{ST} results. In particular, the highest PBS values were present in: *SLC40A1* (rs1439816: AFR=0.27, EUR=0.05, SAS=0.0) and *TMPRSS6* (rs855791: AFR=0.16, EUR=0.0, SAS=0.07; rs855788: AFR=0.29, EUR=0, SAS=0.04; rs5756506: AFR=0.25, EUR=0.0, SAS=0.07) (**S3 Table**). These values are above the top five-percentile threshold of genome-wide PBS values (PBS> 0.156) described by others [41,42].

Finally, we investigated if any signals of recent positive selection could be detected in these genes by using the Integrated Haplotype Score (iHS) values from the Haplotter and HGDP selection browsers. The iHS statistic is based on the LD surrounding a positively selected allele compared with the LD around the alternative variant in the same position [46]. A positive iHS score (iHS > 2) means that the haplotypes on the ancestral allele background are longer than those with the derived allele [46]. A negative iHS score (iHS < -2) means that the haplotypes on the derived allele background are longer and are under selection. No clear evidence of selection was shown in the genomic regions containing *HAMP, TMPRSS6, TF and TRF2* (iHS<1). However, values of iHS scores close to 2 were found for the regions containing *SLC40A1* (e.g. rs1439816: iHS =1.8 (East Asian-Hapmap ASN), iHS=2 (European HapMap CEU) and *HFE* (e.g. rs198846: iHS=1.8 ASN), suggesting a high frequency of longer haplotypes with the ancestral allele. *Other studies have* suggested

that the *HEF* locus could be under positive selection in both European and Asian populations [75].

Discussion

In this study we identified a significant lack of data on the genetic influences of iron status in African populations. This finding highlights a critical gap since African populations have high genetic diversity, and information from other populations may not be transferable to Africans [76,77]. African-specific studies on the genetic influences of iron status will help increase our understanding of the role played by genetic risk factors in the prevalence of anaemia in sub-Saharan Africa.

We used genotype data of populations from the Keneba Biobank at MRCG at LSHTM, The Gambia [33] and the 1000 Genome project [34] to describe the minor allele frequencies and differences in risk alleles in SNPs associated with iron imbalances or iron biomarkers. The allele frequencies of the available SNPs from the Gambian participants in the Keneba Biobank population were very similar to the Gambian population in the 1000 Genomes project. Both the Keneba Biobank population and 1000 Genomes Project included Gambians from the same ethnic group the Mandinka [33,34], which is the largest ethnic group in The Gambia. However, several other ethnic groups live in The Gambia, including Fula and Wolof ethic groups [78]. Variability in disease risk and nutrition status between the Fula and the Mandinka ethnic groups has been reported [79]. This finding is consistent with the inter-population genetic variability within African populations, which may also influence differences in disease susceptibility. Thus, future work could investigate the genetic

diversity in the genes related to iron imbalances in non-Mandinka ethnic groups in The Gambia to determine their possible effect on impaired iron status.

Substantial differences in minor allele frequencies were observed when comparing the African versus non-African populations. The major differences occur in SNPs in *SLC40A1* and *TMPRSS6* genes. *SLC40A1* encodes ferroportin, a transmembrane transport protein which is the only known mammalian iron exporter [80]. The *SLC40A1* Q248H variant (rs11568350) is rare globally except in populations of African ancestry populations, where it reaches frequencies of ~5% [34]. The Q248H variant is associated with increased serum ferritin, decreased hepcidin concentrations and the risk of iron-loading in African populations [57,81]. Also, *SLC40A1* Q248H is associated with modest protection against anaemia and iron deficiency in African children [51,82]. We found significant differences in allelic frequencies for variants in the *TMPRSS6* gene which encodes for Matriptase-2, a type II transmembrane serine protease that negatively regulates hepcidin synthesis [23,83]. Impaired matriptase-2 activity leads to inappropriately raised hepcidin levels [84,85], which results in restricted iron absorption and release from storage sites [17]. Several SNPs in *TMPRSS6* had allele frequencies that are significantly different between African and non-African populations. These variants include rs855791, which has a low MAF (<10%) in African populations and reaches more than 35% in other populations. *TMPRSS6* rs855791 is associated with iron deficiency anaemia and IRIDA, with elevated hepcidin, reduced iron and reduced haemoglobin indices [20,21,84,86,87]. Differences in allele frequencies between continents have been described in many other genetic markers across the genome using data from the 1000 Genomes project

[88,89] Therefore, the observed large allele frequency differences in SNPs associated with iron differences could be the result of demographic differences.

To understand if the differences in the observed allele frequencies could lead to differences in over- or under-representation of risk alleles leading to iron imbalances, we explored the frequencies of the combined risk alleles across the genes. We found that African populations, including the Gambian population from the Keneba Biobank, had a significantly lower number of alleles associated with the risk of anaemia or low iron. Similarly, we observed a lower number of risk alleles associated with high iron, or iron overload in Africans. This observation is likely because most of the studies were conducted in non-African populations. However, it is also possible that these differences are due to natural selection processes to balance the environmental risk factors to which African populations are exposed. For example, malnutrition and infections (e.g. helminths and malaria parasites) can lead to anaemia or limit iron overload which can increase susceptibility to certain infections (e.g bacterial). It is possible that the allele frequency differences between populations we described have occurred through founder effects as humans migrated out of Africa rather than through selective pressure. Possible signals of selection have only been observed for one SNP in *SLC40A1* and three SNPS in *TMPRSS6*, which have the highest F_{ST} and PBS values in Africa.

Our study has limitations. These include the potential for bias in the SNPs selection from the literature as there is an overrepresentation of studies related to genetics of iron imbalances in European and Asian populations. Also, it was difficult to ascertain the risk allele for several variants either because they were not described by the

original study and/or the different studies used different genotyping platforms. In addition, although some risk alleles have been confirmed in more than one ethnic group (46% of the SNPs), for other SNPs it is possible that the alleles have different effects across populations and this could affect the risk allele analysis. Overall, our study highlights a major gap in genetic studies in Africa and the need to perform genetic studies in African populations.

We also observed a lower linkage disequilibrium between SNPs in African populations. For example, the *TMPRSS6* rs4820268 is in strong LD with *TMPRSS6* rs855791 in Europeans [90], but we found that these two SNPs are in weak LD in the Keneba Biobank population. This should be taken into account when performing association studies and selecting tag SNPs. In this setting, it may be easier to fine-scale map “causal” variants, but more difficult to identify the novel putative loci in a GWAS. Also, as iron imbalances can be due to multiple factors, it is critical to complement genetic studies with detailed meta-data collection, including detailed nutritional status, iron biomarkers, and clinical histories. Alternatively the effects of the variants can be studied prospectively using recall-by-genotype methods [91] that can also interrogate the dynamic responses to, for instance, the administration of iron supplements. Follow-up GWAS and candidate gene studies will be important to understand the genetic underpinning the geographic variation in the prevalence of iron imbalances disorders.

In conclusion, this study identified a substantial disparity in allele frequencies of genetic variants associated with iron, between Africans and other populations. We also, identified the scarcity of data on the genetic influences of iron status in Africa.

Given the high burden of iron deficiency in sub-Saharan Africa, particularly in child-bearing women and children, comprehensive mapping of the genetic influences on iron status may help lay the foundation for future studies and assist in developing future iron intervention strategies.

Supporting information

The 1000 Genomes data is publicly available (www.internationalgenome.org). The Keneba Biobank genotyping data for the 13 SNPs used in this study is available in **Table S4.**

Supporting information captions

Fig S1: Linkage disequilibrium (LD) plots in SNPs in *HAMP, SLC40A1*, *TMPRSS6*, *TF*, *HFE* and *TFR2* genes. LD plot showing D prime values in SNPs associated with iron imbalances in (A) African populations, (B) European populations and (C) Gambian population in the Keneba Biobank.

S1 Table: Details of the fifty SNPs identified in the six genes that are associated with iron imbalance

S2 Table: Details of populations where each SNP was reported and the associated phenotypes

S3 Table: Population Branch Statistic (PBS) values involving the comparison of three populations.

S4 Table: Genotyping data for 13 SNPs for the Gambian population in the Keneba Biobank

Acknowledgements

The Authors wish to thank the staff of MRCG at LSHTM Keneba Field Station for contribution to the Keneba Biobank sample collection and Dr. Branwen J Hennig for leading the Biobank sample collection. The Population of West Kiang District, The Gambia for participating in the Biobank project. We also thank Ms. K Pearce at the Core Genomics of the Institute of Child Health, UCL, and our colleagues of the Mal-ED consortium at the Centre for Public Health Genomic, University of Virginia, USA, in particular Drs J Mychaleckyj and U Nayak, for their involvement in the generation of the genetic data. We also thank Dr Neneh Sallah, LSHTM for helping to review the SNPs selection Table.

References

1. Camaschella C. Iron deficiency. Blood. 2019;133: 30–39. doi:10.1182/blood-2018-05-815944
2. WHO. Global Burden of Disease Study 2017. Lancet. 2017; 1–7.
3. Kassebaum NJ. The Global Burden of Anemia. Hematol Oncol Clin North Am. Elsevier Inc; 2016;30: 247–308. doi:10.1016/j.hoc.2015.11.002
4. Khaskheli M-N, Baloch S, Sheeba A, Baloch S, Khaskheli FK. Iron deficiency anaemia is still a major killer of pregnant women. Pakistan J Med Sci. 2016;32: 630–4. doi:10.12669/pjms.323.9557
5. Teshome EM, Andang PEA, Osoti V, Terwel SR, Otieno W, Demir Y, et al. Daily home fortification with iron as ferrous fumarate versus NaFeEDTA : a randomised , Kenyan children. 2017; 1–16. doi:10.1186/s12916-017-0839-z
6. Pena-Rosas JP, De-Regil LM, Dowswell T, Viteri FE. Intermittent oral iron supplementation during pregnancy. Cochrane database Syst Rev. 2012;7: CD009997. doi:10.1002/14651858.CD009997.pub2.www.cochranelibrary.com
7. Petry N, Jallow B, Sawo Y, Darboe MK, Barrow S, Sarr A, et al. Micronutrient Deficiencies , Nutritional Status and the Age and Non-Pregnant Women of Reproductive Age.
8. Heeney MM, Finberg KE. Iron-Refractory Iron Deficiency Anemia (IRIDA). Hematol Oncol Clin North Am. 2014;28: 637–652. doi:10.1016/j.hoc.2014.04.009
9. Mleczko-Sanecka K, Roche F, Da Silva AR, Call D, D'Alessio F, Ragab A, et al. Unbiased RNAi screen for hepcidin regulators links hepcidin suppression to proliferative Ras/RAF and nutrient-dependent mTOR signaling. Blood. 2014;123: 1574–1585. doi:10.1182/blood-2013-07-515957

10. Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Ward DM, et al. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. Science. 2004;306: 2090–3. doi:10.1126/science.1104742
11. Du X, She E, Gelbart T, Truksa J, Lee P, Xia Y, et al. The serine protease TMPRSS6 is required to sense iron deficiency. Science. 2008;320: 1088–92. doi:10.1126/science.1157121
12. Gozzelino R, Arosio P. Iron homeostasis in health and disease. Int J Mol Sci. 2016;17: 2–14. doi:10.3390/ijms17010130
13. McLaren CE, Garner CP, Constantine CC, McLachlan S, Vulpe CD, Snively BM, et al. Genome-wide association study identifies genetic loci associated with iron deficiency. PLoS One. 2011;6. doi:10.1371/journal.pone.0017390
14. Athiyarath R, Shaktivel K, Abraham V, Singh D, Bondu JD, Chapla A, et al. Association of genetic variants with response to iron supplements in pregnancy. Genes Nutr. Springer Berlin Heidelberg; 2015;10: 25. doi:10.1007/s12263-015-0474-2
15. Valenti L, Rametta R, Dongiovanni P, Motta BM, Canavesi E, Pelusi S, et al. The A736V TMPRSS6 Polymorphism Influences Hepatic Iron Overload in Nonalcoholic Fatty Liver Disease. PLoS One. 2012;7. doi:10.1371/journal.pone.0048804
16. Lo KS, Wilson JG, Lange LA, Folsom AR, Galarneau G, Ganesh SK, et al. Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. Hum Genet. 2011;129: 307–17. doi:10.1007/s00439-010-0925-1
17. Ganz T. Systemic iron homeostasis. 2013; 1721–1741. doi:10.1152/physrev.00008.2013

18. Soranzo N, Spector TD, Mangino M, Kühnel B, Rendon A, Teumer A, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet. Nature Publishing Group; 2009;41: 1182–1190. doi:10.1038/ng.467

19. Ganesh SK, Zakai N a, van Rooij FJ a, Soranzo N, Smith A V, Nalls M a, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet. Nature Publishing Group; 2009;41: 1191–1198. doi:10.1038/ng.466

20. Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, et al. Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. Nat Genet. Nature Publishing Group; 2009;41: 1170–2. doi:10.1038/ng.462

21. Bhatia P, Singh A, Hegde A, Jain R, Bansal D. Systematic evaluation of paediatric cohort with iron refractory iron deficiency anaemia (IRIDA) phenotype reveals multiple TMPRSS6 gene variations. Br J Haematol. 2017;177: 311–318. doi:10.1111/bjh.14554

22. Finberg KE, Whittlesey RL, Fleming MD, Andrews NC, Dc W. Down-regulation of Bmp / Smad signaling by Tmprss6 is required for maintenance of systemic iron homeostasis Down-regulation of Bmp / Smad signaling by Tmprss6 is required for maintenance of systemic iron homeostasis. 2012;115: 3817–3826. doi:10.1182/blood-2009-05-224808

23. De Falco L, Sanchez M, Silvestri L, Kannengiesser C, Muckenthaler MU, Iolascon A, et al. Iron refractory iron deficiency anemia. Haematologica. 2013;98: 845–853. doi:10.3324/haematol.2012.075515

24. Blanco-Rojo R, Baeza-Richer C, López-Parra AM, Pérez-Granados AM, Brichs A, Bertoncini S, et al. Four variants in transferrin and HFE genes as potential

markers of iron deficiency anaemia risk: an association study in menstruating women. Nutr Metab (Lond). 2011;8: 69. doi:10.1186/1743-7075-8-69

25. Lee PL, Halloran C, Trevino R, Felitti V, Beutler E, Scripps T. Human transferrin G277S mutation : a risk factor for iron deficiency anaemia. 2001;

26. Benyamin B, McRae AF, Zhu G, Gordon S, Henders AK, Palotie A, et al. Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels. Am J Hum Genet. The American Society of Human Genetics; 2009;84: 60–5. doi:10.1016/j.ajhg.2008.11.011

27. Barrios M, Moreno-Carralero M-I, Cuadrado-Grande N, Baro M, Vivanco J-L, Morán-Jiménez M-J. The homozygous mutation G75R in the human SLC11A2 gene leads to microcytic anaemia and iron overload. Br J Haematol. 2012;157: 514–6. doi:10.1111/j.1365-2141.2012.09043.x

28. Kloss-Brandstätter A, Erhart G, Lamina C, Meister B, Haun M, Coassin S, et al. Candidate gene sequencing of SLC11A2 and TMPRSS6 in a family with severe anaemia: Common SNPs, rare haplotypes, no causative mutation. PLoS One. 2012;7: 1–8. doi:10.1371/journal.pone.0035015

29. Benyamin B, Ferreira MAR, Willemsen G, Gordon S, Middelberg RPS, McEvoy BP, et al. Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. Nat Genet. 2009;41: 1173–5. doi:10.1038/ng.456

30. Gichohi-Wainaina WN, Tanaka T, Towers GW, Verhoef H, Veenemans J, Talsma EF, et al. Associations between Common Variants in Iron-Related Genes with Haematological Traits in Populations of African Ancestry. Samuels DC, editor. PLoS One. 2016;11: e0157996. doi:10.1371/journal.pone.0157996

31. Hunt SE, Mclaren W, Gil L, Thormann A, Schuilenburg H, Sheppard D, et al.

Ensembl variation resources. 2018; 1–12. doi:10.1093/database/bay119

32. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29: 308–311. doi:10.1093/nar/29.1.308

33. Hennig BJ, Unger SA, Dondeh BL, Hassan J, Hawkesworth S, Jarjou L, et al. Cohort profile: The Kiang West Longitudinal Population Study (KWLPS)-a platform for integrated research and health care provision in rural Gambia. Int J Epidemiol. 2017;46: 1–12. doi:10.1093/ije/dyv206

34. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526: 68–74. doi:10.1038/nature15393

35. Warnes G, Gorjanc G, Leisch F, Man M. Package ‘ genetics .’ 2019;

36. R Core Team. A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.

37. Jombart T, Kamvar ZN, Collins C, Lustrik R, Beugin M-P, Knaus BJ, et al. Package “adegenet” Encoding UTF-8 Title Exploratory Analysis of Genetic and Genomic Data. 2018;

38. Goudet J, Jombart T. Estimating and Tests of Hierarchical F-Statistics. Evolution (N Y). 1998;52: 950. doi:10.2307/2411227

39. Paradis E, Jombart T, Kamvar ZN, Knaus B, Schliep K, Potts A, et al. Population and Evolutionary Genetics Analysis System. 2019;

40. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science (80-). American Association for the Advancement of Science; 2010;329: 75–78. doi:10.1126/science.1190371

41. Gouveia MH, Bergen AW, Borda V, Nunes K, Leal TP, Ogwang MD, et al.

Genetic signatures of gene flow and malariadriven natural selection in Sub-Saharan populations of the “endemic burkitt lymphoma belt.” PLoS Genet. Public Library of Science; 2019;15: e1008027. doi:10.1371/journal.pgen.1008027

42. Crawford JE, Amaru R, Song J, Julian CG, Racimo F, Cheng JY, et al. Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. Am J Hum Genet. Cell Press; 2017;101: 752–767. doi:10.1016/j.ajhg.2017.09.023
43. Ding K, Kullo IJ. Geographic differences in allele frequencies of susceptibility SNPs for cardiovascular disease. BMC Med Genet. BioMed Central; 2011;12: 55. doi:10.1186/1471-2350-12-55
44. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science (80-). Science; 2008;319: 1100–1104. doi:10.1126/science.1153717
45. Myles S, Davison D, Barrett J, Stoneking M, Timpson N. Worldwide population differentiation at disease-associated SNPs. BMC Med Genomics. Springer Nature; 2008;1: 22. doi:10.1186/1755-8794-1-22
46. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A Map of Recent Positive Selection in the Human Genome. Hurst L, editor. PLoS Biol. Public Library of Science; 2006;4: e72. doi:10.1371/journal.pbio.0040072
47. Pybus M, Olio GMD, Luisi P, Uzkudun M, Pavlidis P, Laayouni H, et al. 1000 Genomes Selection Browser 1 . 0 : a genome browser dedicated to signatures of natural selection in modern humans. 2014;42: 903–909. doi:10.1093/nar/gkt1188

48. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage [Internet]. Science. 2006. pp. 1614–1620. doi:10.1126/science.1124309

49. Danquah I, Gahutu J-B, Zeile I, Musemakweri A, Mockenhaupt FP. Anaemia, iron deficiency and a common polymorphism of iron-regulation, TMPRSS6 rs855791, in Rwandan children. Trop Med Int Health. 2014;19: 117–22. doi:10.1111/tmi.12216

50. Masaisa F, Breman C, Gahutu JB, Mukiibi J, Delanghe J, Philippé J. Ferroportin (SLC40A1) Q248H mutation is associated with lower circulating serum hepcidin levels in Rwandese HIV-positive women. Ann Hematol. 2012;91: 911–916. doi:10.1007/s00277-011-1400-3

51. Kasvosve I, Gomo ZAR, Nathoo KJ, Matibe P, Mudenge B, Loyevsky M, et al. Effect of ferroportin Q248H polymorphism on iron status in. 2018; 1102–1106.

52. Gichohi-Wainaina, W. N. Melse-Boonstra, A. Swinkels, D. W. Zimmermann, M. B. Feskens, E. J. Towers GW. Common variants and haplotypes in the TF, TNF- alpha , and TMPRSS6 genes are associated with iron status in a female black South. J Nutr 2015. 2015;145: 945–953. doi:10.3945/jn.114.209148.945

53. McLaren CE, McLachlan S, Garner CP, Vulpe CD, Gordeuk VR, Eckfeldt JH, et al. Associations between single nucleotide polymorphisms in iron-related genes and iron status in multiethnic populations. PLoS One. 2012;7. doi:10.1371/journal.pone.0038339

54. Pichler I, Minelli C, Sanna S, Tanaka T, Schwienbacher C, Naitza S, et al. Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels. Hum Mol Genet. 2011;20: 1232–40. doi:10.1093/hmg/ddq552

55. Radio FC, Majore S, Aurizi C, Sorge F, Biolcati G, Bernabini S, et al. Hereditary hemochromatosis type 1 phenotype modifiers in Italian patients. The controversial role of variants in HAMP, BMP2, FTL and SLC40A1 genes. Blood Cells Mol Dis. Elsevier B.V.; 2015;55: 71–5. doi:10.1016/j.bcmd.2015.04.001

56. Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. Nat Commun. 2014;5: 4926. doi:10.1038/ncomms5926

57. Rivers CA, Barton JC, Gordeuk VR, Acton RT, Speechley MR, Snively BM, et al. Association of ferroportin Q248H polymorphism with elevated levels of serum ferritin in African Americans in the Hemochromatosis and Iron Overload Screening (HEIRS) Study. Blood Cells, Mol Dis. 2007;38: 247–252. doi:10.1016/j.bcmd.2006.12.002

58. Javaheri-Kermani M, Farazmandfar T, Ajami A, Yazdani Y. Impact of hepcidin antimicrobial peptide on iron overload in tuberculosis patients. Scand J Infect Dis. 2014;46: 693–696. doi:10.3109/00365548.2014.929736

59. Andreani M, Radio FC, Testi M, De Bernardo C, Troiano M, Majore S, et al. Association of hepcidin promoter c.-582 A>G variant and iron overload in thalassemia major. Haematologica. 2009;94: 1293–1296. doi:10.3324/haematol.2009.006270

60. Jackson HA, Carter K, Darke C, Guttridge MG, Ravine D, Hutton RD, et al. HFE mutations, iron deficiency and overload in 10 500 blood donors. Br J Haematol. 2001;114: 474–484. doi:10.1046/j.1365-2141.2001.02949.x

61. An P, Wu Q, Wang H, Guan Y, Mu M, Liao Y, et al. TMPRSS6, but not TF, TFR2 or BMP2 variants are associated with increased risk of iron-deficiency

anemia. Hum Mol Genet. 2012;21: 2124–2131. doi:10.1093/hmg/dds028

62. Pichler I, Minelli C, Sanna S, Tanaka T, Schwienbacher C, Naitza S, et al. Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels. Hum Mol Genet. Oxford University Press; 2011;20: 1232–40. doi:10.1093/hmg/ddq552

63. Piao W, Wang L, Zhang T, Wang Z, Shangguan S, Sun J, et al. A single-nucleotide polymorphism in transferrin is associated with soluble transferrin receptor in Chinese adolescents. Asia Pac J Clin Nutr. 2017;26: 1170–1178. doi:10.6133/apjcn.112016.04

64. Li J, Lange LA, Duan Q, Lu Y, Singleton AB, Zonderman AB, et al. Genome-wide admixture and association study of serum iron, ferritin, transferrin saturation and total iron binding capacity in African Americans. Hum Mol Genet. 2015;24: 572–581. doi:10.1093/hmg/ddu454

65. Beutler E, Felitti V, Gelbart T, Waalen J. Haematological effects of the C282Y HFE mutation in homozygous and heterozygous states among subjects of northern and southern European ancestry. Br J Haematol. 2003;120: 887–893. doi:10.1046/j.1365-2141.2003.04215.x

66. Galesloot TE, Geurts-Moespot AJ, den Heijer M, Sweep FCGJ, Fleming RE, Kiemeney L a LM, et al. Associations of common variants in HFE and TMPRSS6 with iron parameters are independent of serum hepcidin in a general population: a replication study. J Med Genet. 2013;50: 593–8. doi:10.1136/jmedgenet-2013-101673

67. De Falco L, Tortora R, Imperatore N, Bruno M, Capasso M, Girelli D, et al. The role of TMPRSS6 and HFE variants in iron deficiency anemia in celiac disease. Am J Hematol. 2018;93: 383–393. doi:10.1002/ajh.24991

68. Pichler I, Del Greco M F, Gögele M, Lill CM, Bertram L, Do CB, et al. Serum iron levels and the risk of Parkinson disease: a Mendelian randomization study. PLoS Med. 2013;10: e1001462. doi:10.1371/journal.pmed.1001462

69. Garewal G, Das R, Ahluwalia J, Marwaha RK. Prevalence of the H63D mutation of the HFE in north India: Its presence does not cause iron overload in beta thalassemia trait. Eur J Haematol. 2005;74: 333–336. doi:10.1111/j.1600-0609.2004.00390.x

70. Sørensen E, Rigas AS, Thørner LW, Burgdorf KS, Pedersen OB, Petersen MS, et al. Genetic factors influencing ferritin levels in 14,126 blood donors: Results from the Danish Blood Donor Study. Transfusion. 2016;56: 622–627. doi:10.1111/trf.13397

71. Kullo IJ, Ding K, Jouni H, Smith CY, Chute CG. A genome-wide association study of red blood cell traits using the Electronic Medical Record. PLoS One. 2010;5: 1–9. doi:10.1371/journal.pone.0013011

72. Gordeuk VR, Lovato L, Barton JC, Mph MV, Mclaren G, Acton RT, et al. Dietary iron intake and serum ferritin concentration in 213 patients homozygous for the HFE C282Y hemochromatosis mutation. 2012;26: 345–349.

73. Chen Z, Tang H, Qayyum R, Schick UM, Nalls MA, Handsaker R, et al. Genome-wide association analysis of red blood cell traits in African Americans: The cogent network. Hum Mol Genet. 2013;22: 2529–2538. doi:10.1093/hmg/ddt087

74. Whitfield JB, Cullen LM, Jazwinska EC, Powell LW, Heath AC, Zhu G, et al. Effects of HFE C282Y and H63D Polymorphisms and Polygenic Background on Iron Stores in a Large Community Sample of Twins. 2000; 1246–1258.

75. Ye K, Cao C, Lin X, Brien KOO, Gu Z. Natural selection on HFE in Asian populations contributes to enhanced non-heme iron absorption. BMC Genet. BMC Genetics; 2015; 1–11. doi:10.1186/s12863-015-0223-y

76. Gomez F, Hirbo J, Tishkoff SA. Genetic variation and adaptation in Africa: implications for human evolution and disease. Cold Spring Harb Perspect Biol. 2014;6: a008524. doi:10.1101/cshperspect.a008524

77. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet. 2008;9: 403–33. doi:10.1146/annurev.genom.9.081307.164258

78. World Population View. Gambia Population 2020 [Internet]. 2020.

79. Wilson AL, Bradley J, Kandeh B, Salami K, D'Alessandro U, Pinder M, et al. Is chronic malnutrition associated with an increase in malaria incidence? A cohort study in children aged under 5 years in rural Gambia. Parasites and Vectors. Parasites & Vectors; 2018;11: 1–11. doi:10.1186/s13071-018-3026-y

80. Lane DJR, Merlot AM, Huang ML-H, Bae D-H, Jansson PJ, Sahni S, et al. Cellular iron uptake, trafficking and metabolism: Key molecules and mechanisms and their roles in disease. Biochim Biophys Acta - Mol Cell Res. 2015;1853: 1130–1144. doi:https://doi.org/10.1016/j.bbamcr.2015.01.021

81. Gordeuk VR, Caleffi A, Corradini E, Ferrara F, Jones RA, Castro O, et al. Iron overload in Africans and African-Americans and a common mutation in the SCL40A1 (ferroportin 1) gene. Blood Cells Mol Dis. 31: 299–304. doi:10.1016/s1079-9796(03)00164-5

82. Muriuki JM, Mentzer AJ, Kimita W, Ndungu FM, MacHaria AW, Webb EL, et al. Iron Status and Associated Malaria Risk among African Children. Clin Infect

Dis. 2019;68: 1807–1814. doi:10.1093/cid/ciy791

83. Wahedi M, Wortham AM, Kleven MD, Zhao N, Jue S, Enns CA, et al. Matriptase-2 suppresses hepcidin expression by cleaving multiple components of the hepcidin induction pathway. J Biol Chem. 2017;292: 18354–18371. doi:10.1074/jbc.M117.801795

84. Nai A, Pagani A, Silvestri L, Campostrini N, Corbella M, Girelli D, et al. TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. Blood. 2011;118: 4459–4462. doi:10.1182/blood-2011-06-364034

85. De Falco L, Silvestri L, Kannengiesser C, Morán E, Oudin C, Rausa M, et al. Functional and clinical impact of novel TMPRSS6 variants in iron-refractory iron-deficiency anemia patients and genotype-phenotype studies. Hum Mutat. 2014;35: 1321–9. doi:10.1002/humu.22632

86. Delbini P, Vaja V, Graziadei G, Duca L, Nava I, Refaldi C, et al. Genetic variability of TMPRSS6 and its association with iron deficiency anaemia. Br J Haematol. 2010;151: 281–284. doi:10.1111/j.1365-2141.2010.08349.x

87. Pei SN, Ma MC, You HL, Fu HC, Kuo CY, Rau KM, et al. TMPRSS6 rs855791 polymorphism influences the susceptibility to iron deficiency anemia in women at reproductive age. Int J Med Sci. 2014;11: 614–619. doi:10.7150/ijms.8582

88. Elhaik E. Empirical Distributions of FST from Large-Scale Human Polymorphism Data. Mailund T, editor. PLoS One. Public Library of Science; 2012;7: e49837. doi:10.1371/journal.pone.0049837

89. Hofer T, Ray N, Wegmann D, Excoffier L. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. Ann Hum Genet. Blackwell

Publishing Ltd; 2009;73: 95–108. doi:10.1111/j.1469-1809.2008.00489.x

90. Yu F, Keinan A, Chen H, Ferland RJ, Hill RS, Mignault AA, et al. Detecting natural selection by empirical comparison to random regions of the genome. Hum Mol Genet. 2009;18: 4853–67. doi:10.1093/hmg/ddp457

91. Corbin LJ, Tan VY, Hughes DA, Wade KH, Paul DS, Tansey KE, et al. Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference. Nat Commun. 2018;9: 711. doi:10.1038/s41467-018-03109-y

Supplemental material

Differences in the frequency of genetic variants associated with iron imbalance among global populations

Momodou W. Jallow[1,2], Carla Cerami[1], Taane G. Clarke[2], Andrew M. Prentice[1] and Susana Campino[2]

S1 Table: Details of the fifty SNPs identified in the six genes that are associated with iron imbalance

SNPs	Loci	Type of variant (amino acid change)	Minor Allele	Major Allele	Risk allele	Effect on serum iron[1]	Minor Allele Frequency										References
							1000 Genomes							HapMap		Keneba Biobank	
							Global (All)	AFR (All)	GWD	EUR	EAS	SAS	AMR	YRI	CEU		
rs10421768	*HAMP*	intron variant	G	A	A	High	0.16	0.19	0.26	0.24	0.03	0.20	0.14	0.21	0.13	NA	(1–4)
rs1799945	*HFE*	missense variant (aa: H/D)	G	C	G	High	0.07	0.01	0.00	0.17	0.03	0.07	0.12	0.01	0.13	0.01	(5–17)
rs1800562	*HFE*	Missense variant (C282Y)	A	G	A	High	0.01	0.00	0.00	0.04	0.00	0.00	0.02	0.00	0.05	0	(13,15,16,18–20)
rs198846	*close to HFE*	Intron variant	A	G	A	High	0.11	0.12	0.05	0.18	0.02	0.07	0.14	0.21	0.16	NA	(21,22)
rs129128	close to *HFE*	Intron variant	C	T	C	High	0.07	0.01	0.00	0.16	0.03	0.09	0.11	0.01	0.14	NA	(23)
rs744653	close to *SLC40A1*	regulatory region variant	C	T	T	Moderates HH[2]	0.14	0.21	0.18	0.16	0.04	0.19	0.08	0.18	0.15	NA	(5)
rs1439816	*SLC40A1*	intron variant	C	G	G	Moderates HH[2]	0.34	0.73	0.74	0.16	0.18	0.25	0.23	0.76	0.17	NA	(4)
rs11568350 (Q248H)	*SLC40A1*	Missense variant (Q248H)	A	C	A	High	0.01	0.05	0.06	0.00	0.00	0.00	0.00	0.05	0.00	0.05	(24–26)
rs2280673	close to *TF*	intron variant RAB6B	A	C	NA	Low	0.49	0.41	0.38	0.37	0.61	0.56	0.47	0.40	0.34	NA	(18)

Table S1 Continued

SNPs	Loci	Type of variant (amino acid change)	Minor Allele	Major Allele	Risk allele	Effect on serum iron[1]	Minor Allele Frequency										References
							1000 Genomes							HapMap		Keneba Biobank	
							Global (All)	AFR (All)	GWD	EUR	EAS	SAS	AMR	YRI	CEU		
rs1867504	*TF*	Intron variant	A	G	A	High	0.42	0.22	0.24	0.49	0.50	0.52	0.43	0.48	0.23	NA	(2)
rs9872999	*TF*	Intron variant	C	T	NA	High	0.33	0.35	0.39	0.47	0.26	0.25	0.30	NA	NA	NA	(12)
rs8177179	*TF*	Intron variant	G	A	A	Moderates HH[2]	0.34	0.36	0.40	0.47	0.26	0.26	0.30	0.28	0.42	NA	(5)
rs1799852	*TF*	Synonymous variant (L247L)	A	G	A	High	0.14	0.05	0.10	0.14	0.22	0.19	0.14	0.06	0.06	0.07	(2,13,18, 20,27)
rs12493168	*TF*		G	A	NA	Low	0.07	0.01	0.00	0.13	0.00	0.08	0.21	0.01	0.17	NA	(27)
rs1799899 (G277S)	*TF*	Missense variant (G277S)	A	G	A	Conflict[3]	0.03	0.00	0.00	0.07	0.00	0.05	0.04	0.00	0.04	NA	(28,29)
rs3811658	*TF*	Intron variant	T	C	T	Conflict[3]	0.32	0.12	0.10	0.35	0.43	0.41	0.39	0.01	0.37	NA	(2,27,30)
rs8177248	*TF*	intron variant	T	C	NA	Low	0.31	0.08	0.07	0.35	0.43	0.41	0.39	0.04	0.36	NA	(30)
rs8177253	*TF*	intron variant	T	C	T	Low	0.35	0.22	0.15	0.35	0.43	0.41	0.40	0.22	0.36	NA	(12)
rs1405023	*TF*	intron variant	C	T	NA	High	0.44	0.62	0.62	0.44	0.33	0.38	0.35	NA	NA	NA	(27)

Table S1 Continued

SNPs	Loci	Type of variant (amino acid change)	Minor Allele	Major Allele	Risk allele	Effect on serum iron[1]	Minor Allele Frequency										References
							1000 Genomes							HapMap		Keneba Biobank	
							Global (All)	AFR (All)	GWD	EUR	EAS	SAS	AMR	YRI	CEU		
rs1880669	*TF*	intron variant	T	C	NA	Conflict[3]	0.50	0.65	0.66	0.39	0.50	0.44	0.44	0.70	0.39	NA	(27,30,31)
rs3811647	*TF*	intron variant	A	G	A	Low	0.34	0.19	0.15	0.35	0.42	0.41	0.39	0.17	0.36	0.14	(6,12,13,18,27, 32–36)
rs1358024	*TF*	intron variant	T	C	NA	Low	0.19	0.01	0.00	0.19	0.39	0.27	0.17	0.00	0.18	NA	(18,27,30,33)
rs1525892	*TF*	intron variant	A	G	A	Conflict[3]	0.36	0.26	0.23	0.35	0.47	0.41	0.40	0.37	0.23	NA	(2,30,33)
rs1049296	*TF*	Missense variant (S589P)	T	C	NA	High	0.16	0.06	0.02	0.14	0.26	0.23	0.12	0.07	0.16	0.01	(27)
rs7638018	*TF*	intron variant	G	A	NA	Low	0.33	0.15	0.14	0.35	0.42	0.41	0.40	0.15	0.36	NA	(30)
rs1830084	*TF*	3 prime UTR variant	T	A	Uncertain risk allele[4]	Low	0.32	0.11	0.08	0.34	0.46	0.40	0.40	0.13	0.35	NA	(12,30)
rs7385804	*TFR2*	Intron variant	C	A	C	Conflict[3]	0.31	0.33	0.30	0.38	0.24	0.32	0.29	0.35	0.38	NA	(2,5,6,31,32,37)
rs2235321	*TMPRSS6*	Synonymous variant (Y739Y)	A	G	A	Low	0.36	0.41	0.44	0.42	0.41	0.26	0.21	0.38	0.39	0.44	(38–40)
rs855791	*TMPRSS6*	Missense variant (A736V)	A	G	A	Low	0.40	0.10	0.10	0.39	0.57	0.54	0.49	0.12	0.41	0.07	(5,8,10,15–17,21,22,32,36, 39,41–59)
rs78174698	*TMPRSS6*	missense variant (P555S)	A	G	NA	Low	0.03	0.01	0.02	0.00	0.01	0.12	0.00	0.02	0.00	0.01	(43)

Table S1 Continued

SNPs	Loci	Type of variant (amino acid change)	Minor Allele	Major Allele	Risk allele	Effect on serum iron[1]	Minor Allele Frequency										References
							1000 Genomes							HapMap		Keneba Biobank	
							Global (All)	AFR (All)	GWD	EU R	EAS	SAS	AMR	YRI	CEU		
rs5756504	*TMPRSS6*	Intron variant	T	C	T	High	0.43	0.67	0.65	0.40	0.42	0.26	0.24	0.71	0.33	NA	(22,56,57,60)
rs5756506	*TMPRSS6*	Intron variant	C	G	C	High	0.47	0.83	0.82	0.40	0.43	0.26	0.26	0.85	0.35	0.84	(27,37,47)
rs4820268	*TMPRSS6*	Missense variant (D521E)	G	A	G	Low	0.46	0.28	0.27	0.42	0.56	0.57	0.53	0.21	0.48	0.27	(2,6,21,22,32,39,42,47,57,61,62)
rs2413450	*TMPRSS6*	Intron variant	T	C	T	Low	0.42	0.12	0.12	0.41	0.56	0.56	0.52	0.12	0.48	0.17	(2,47,53,63)
rs2072860	*TMPRSS6*	Intron variant	G	A	NA	Conflict[3]	0.46	0.28	0.27	0.42	0.57	0.57	0.53	NA	NA	NA	(12,43)
rs9610643	*TMPRSS6*	Intron variant	A	G	NA	Low	0.38	0.60	0.59	0.33	0.40	0.23	0.22	NA	NA	NA	(43)
rs855788	*TMPRSS6*	intron variant	A	G	NA	High	0.49	0.90	0.86	0.44	0.30	0.35	0.27	0.95	0.31	NA	(57)
rs2543519	*TMPRSS6*	Intron variant	G	A	NA	Low	0.25	0.40	0.43	0.21	0.17	0.25	0.14	0.36	0.21	NA	(39,43)
rs2111833	*TMPRSS6*	Synonymous variant (S>S)	T	C	T	Conflict[3]	0.31	0.38	0.31	0.39	0.31	0.24	0.20	0.42	0.34	NA	(4) (30)
rs2235324	*TMPRSS6*	Missense variant (K253E)	G	A	G	Low	0.39	0.40	0.43	0.43	0.40	0.37	0.33	0.43	0.35	0.45	(38–40,47,51,57)

Table S1 Continued

SNPs	Loci	Type of variant (amino acid change)	Minor Allele	Major Allele	Risk allele	Effect on serum iron[1]	Minor Allele Frequency										References
							1000 Genomes							HapMap		Keneba Biobank	
							Global (All)	AFR (All)	GWD	EUR	EAS	SAS	AMR	YRI	CEU		
rs1421312	*TMPRSS6*	intron variant	G	A	NA	High	0.47	0.60	0.58	0.42	0.40	0.50	0.35	0.62	0.47	NA	(30,57)
rs5756512	*TMPRSS6*	intron variant	T	C	NA	Low	0.33	0.28	0.26	0.42	0.33	0.36	0.23	NA	NA	NA	(43)
rs2160906	*TMPRSS6*	Intron variant	A	G	NA	Low	0.13	0.06	0.05	0.19	0.18	0.14	0.12	0.05	0.20	NA	(36)
rs732756	*TMPRSS6*	Intron variant	C	T	NA	Low	0.14	0.08	0.08	0.19	0.18	0.14	0.12	0.06	0.20	NA	(43)
rs228904	*TMPRSS6*	Intron variant	G	A	NA	High	0.14	0.08	0.08	0.19	0.18	0.14	0.12	0.06	0.20	NA	(57)
rs11704654	*TMPRSS6*	Synonymous variant (P33P)	T	C	NA	Low	0.15	0.15	0.14	0.19	0.13	0.16	0.11	0.16	0.25	NA	(39,42)
rs5756516	*TMPRSS6*	Intron variant	T	C	NA	Low	0.32	0.30	0.27	0.42	0.33	0.20	0.35	0.31	0.43	NA	(43)
rs228916	*TMPRSS6*	5 prime UTR variant	C	T	T	Low	0.07	0.03	0.00	0.11	0.00	0.09	0.18	0.03	0.08	NA	(5)
rs228921	*TMPRSS6* 2kb Upstream Variant	Intergenic variant	G	A	G	Low	0.41	0.40	0.40	0.41	0.43	0.48	0.31	NA	0.40	NA	(21,35)
rs228918	*TMPRSS6:* 2kb Upstream	Intergenic variant-	A	G	G	Low	0.41	0.40	0.40	0.41	0.43	0.49	0.31	0.34	0.47	NA	(2,21)

[1] The documented effect of each SNP on iron status, based on its influence on iron biomarkers. High: Indicates SNPs that have been associated with elevated iron status as shown by at least iron biomarker signifying elevated iron status. Low indicates SNPs associated with decreased iron status, determined by at least one biomarker signifying low iron.
[2] The only information available about this SNPs is that they modulate hemochromatosis.
[3] We found contradictory information about the effects of these SNPs on iron status. Different papers reported direction of effects of these SNPs on iron status.
NA indicates SNPs that we could not establish the risk allele because it was not stated by the respective studies that reported the SNPs. In the Keneba Biobank, NA indicates SNPs whose genotype data was not present in the Biobank population.
[4]Indicates a SNP in which the effect of the risk allele has not been clearly described in the paper it was reported.
[5]The phenotype associated with the risk allele

AFR, Africans; AMR, Americans; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; EAS, East Asians; EUR, Europeans; GWD, Gambians from Western Division; HAMP, hepcidin antimicrobial peptide; Hb, haemoglobin; HCT, haematocrit; HFE, High fe; HH, hereditary hemochromatosis; IDA, iron deficiency anaemia; MCH, mean corpuscular haemoglobin; NA, not available; SAS, South Asians; SI, serum iron; *SLC40A1*, solute carrier family 40 member 1; SNP, single nucleotide polymorphism; sTfR, soluble transferrin receptor; *TF*, transferrin; *TMPRSS6*, transmembrane protease serine 6; UTR, untranslated region; YRI, Yoruba in Nigeria.

Table S2. Details of populations where each SNP was reported and the associated phenotypes

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs10421768	*HAMP*	G	A	A	Javaheri-Kermani et al., 2014	Iran	Associated with elevated serum iron and low hepcidin levels	(3)
rs10421768	*HAMP*	G	A	A	Andreani et al., 2009	Italy	Associated with elevated liver iron concentration and raised serum ferritin levels	(1)
rs10421768	*HAMP*	G	A	A	Radio et al., 2016	Italy	Homozygotes and heterozygotes has signicficantly reduced transferrin levels	(4)
rs10421768	*HAMP*	G	A	A	Gichohi-Wainaina et al., 2016	Kenyans, Tanzanians, S. Africans and African-Americans	Significant increased in Hb in Kenyans only	(2)
rs1049296	*TF*	T	C	NA	Constantine et al., 2009	European ancestry	Reduced TSAT and serum transferrin	(27)
rs11568350 (Q248H)	*SLC40A1*	A	C	A	Masaisa et al., 2012	Rwanda	Associated with low hepcidin and transferrin, and elevated serum ferritin	(25)
rs11568350 (Q248H)	*SLC40A1*	A	C	A	Kasvosve et al., 2018	Zimbabwe	Associated with elevated ferritin levels and protection against IDA	(24)
rs11568350 (Q248H)	*SLC40A1*	A	C	A	Rivers et al., 2007	African-Americans	Associated with elevated serum ferritin in men	(26)
rs11704654	*TMPRSS6*	T	C	NA	Delbini et al., 2010	Italy	Associated with iron deficiency	(39)
rs11704654	*TMPRSS6*	T	C	NA	Kloss-Brandstatter et al., 2012	Netherlands	Associated with increased serum iron and ferritin	(42)
rs12493168	*TF*	G	A	NA	Constantine et al., 2009	European ancestry	Elevated serum transferritin	(27)

Table S2 continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs129128	*HFE*	C	T	C	Li et al., 2015	USA	Elevated serum iron	(12)
rs1358024	*TF*	T	C	NA	Constantine et al., 2009	European ancestry	Elevated serum transferritin	(27)
rs1358024	*TF*	T	C	NA	Benyamin et al., 2009	Aurtralians	Affects serum transferrin concentrations, but direction of effect was not stated	(18)
rs1405023	*TF*	C	T	NA	Constantine et al., 2009	European ancestry	Reduced serum transferrin levels	(27)
rs1421312	*TMPRSS6*	G	A	NA	McLaren et al., 2012	Whites, African-Americans, Hispanics and Asians (US & Canada)	Increased serum iron and TSAT, and decreased sTfR in Whites	(30)
rs1421312	*TMPRSS6*	G	A	NA	Tanaka et al., 2010	Italy and USA	Significantly associated with reduced iron status	(57)
rs1439816	*SLC40A1*	C	G	G	Radio et al., 2016	Italy	Moderates hereditary hemochromatosis	(4)
rs1525892	*TF*	A	G	A	Gichohi-Wainaina et al., 2016	Kenyans, Tanzanians, S. Africans and African-Americans	Marginally significant higher ferritin concentrations	(2)
rs1525892	*TF*	A	G	A	McLaren et al., 2012	Whites, African-Americans, Hispanics and Asians (US & Canada)	Elevated TIBC in all the populations	(30)
rs1799852	*TF*	A	G	A	Gichohi-Wainaina et al., 2016	South Africa	The A allele is associated with lower serum ferritin	(2)

Table S2 continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs1799852	*TF*	A	G	A	Benyamin, et al., 2009	Australian of European ancestry	Decreased transferrin, increased serum iron, ferritin and TSAT	(18)
rs1799852	*TF*	A	G	A	Blanco-Rojo et al., 2011	Spain	Reduced serum transferrin levels	(13)
rs1799852	*TF*	A	G	A	Constantine et al., 2009	European ancestry	Reduced serum transferrin levels	(27)
rs1799899 (G277S)	*TF*	A	G	A	Sarria et al., 2007	Spain	No significant differences in iron biomarkers between genotypes	(28)
rs1799899 (G277S)	*HFE*	A	G	A	Lee et al., 2001	European ancestry	The variant allele predispose to iron deficiency	(29)
rs1799945 (H63D)	*HFE*	G	C	G	Pichler et al., 2011	Italy and USA	Reduced serum iron	(6)
rs1799945 (H63D)	*HFE*	G	C	G	Athiyarath et al., 2015	India	The G allele was significantly associated with adequate response to iron supplementation	(10)
rs1799945 (H63D)	*HFE*	G	C	G	Blanco-Rojo et al., 2011	Spain	Reduced serum transferrin levels	(13)
rs1799945 (H63D)	*HFE*	G	C	G	Galesloot et al., 2013	Netherlands	Significantly associated with iron and TSAT	(15)
rs1799945 (H63D)	*HFE*	G	C	G	De Falco et al., 2018	Italy	associated with elevated Hb, MCV, serum iron and ferritin levels	(16)
rs1799945 (H63D)	*HFE*	G	C	G	Mast et al., 2012	Multicenter: USA	Increased iron stores	(9)
rs1799945 (H63D)	*HFE*	G	C	G	Sørensen et al., 2015	Denmark	The C alleles are associated with iron deficiency in women	(8)

Table S2 continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs1799945 (H63D)	*HFE*	G	C	G	Whitfield et al., 2000	Australia	Associated with high iron stores	(11)
rs1799945 (H63D)	*HFE*	G	C	G	Benyamin et al., 2014	European ancestry	Reduced serum iron, TSAT and ferritin, and elevated transferrin	(5)
rs1799945 (H63D)	*HFE*	G	C	G	Li et al., 2015	USA	Elevated serum iron	(12)
rs1799945 (H63D)	*HFE*	G	C	G	Pichler et al., 2013	European Ancestry populations	Increased iron stores	(17)
rs1799945 (H63D)	*HFE*	G	C	G	Blanco-Rojo et al., 2014	European ancestry	Associated with protection against iron deficiency	(20)
rs1799945 (H63D)	*HFE*	G	C	G	Garewal et al., 2005	India	No effect on iron status	(7)
rs1799945 (H63D)	*HFE*	G	C	G	Beutler et al., 2003	European ancestry	Elevated Hb, TSAT, ferritin and lower anaemia prevalence	(14)
rs1799945 (H63D)	*HFE*	G	C	G	Jackson et al., 2001	UK: Wales	Elevated serum ferritin, TSAT and Hb, and reduced UIBC	(64)
rs1800562 (C282Y)	*HFE*	A	G	A	Benyamin, et al., 2009	Australian of European ancestry	Increased serum iron, ferritin and TSAT, decreased transferrin	(18)
rs1800562 (C282Y)	*HFE*	A	G	A	Kullo et al., 2010	USA	Elevated MCH	(22)
rs1800562 (C282Y)	*HFE*	A	G	A	Blanco-Rojo et al., 2011	Spain	Reduced serum transferrin levels	(13)
rs1800562 (C282Y)	*HFE*	A	G	A	Galesloot et al., 2013	Netherlands	Significantly associated with ferritin, iron, TSAT and TIBC	(15)
rs1800562 (C282Y)	*HFE*	A	G	A	Benyamin et al., 2009	Aurtralians	Reduced transferrin, raised serum iron, TSAT, ferritin, Hb and MCV	(36)

Table S2 continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs1800562 (C282Y)	*HFE*	A	G	A	Seiki et al., 2018	Japanese	Elevated MCV, Hb	(60)
rs1800562 (C282Y)	*HFE*	A	G	A	Gordeuk et al., 2017	Multi-ethnic: USA and Canada	Associated with elevated ferritin levels	(19)
rs1800562 (C282Y)	*HFE*	A	G	A	De Falco et al., 2018	Italians	Carriage of C282Y was higher in celiac disease cases in than in controls.	(16)
rs1800562 (C282Y)	*HFE*	A	G	A	Sørensen et al., 2015	Denmark	The G allele is associated with lower iron stores	(8)
rs1800562 (C282Y)	*HFE*	A	G	A	Whitfield et al., 2000	Australians	Associated with high iron stores	(11)
rs1800562 (C282Y)	*HFE*	A	G	A	Benyamin et al., 2014	European ancestry	Elevated iron, TSAT and ferritin, and reduced tranferrin	(5)
rs1800562 (C282Y)	*HFE*	A	G	A	Li et al., 2015	USA	Elevated ferritin and low TIBC	(12)
rs1800562 (C282Y)	*HFE*	A	G	A	Pichler et al., 2013	European Ancestry population	Increased iron stores	(17)
rs1800562 (C282Y)	*HFE*	A	G	A	Traglia et al., 2011	Italians	Elevated ferritin and decreased hepcidi/ferritin ration	(45)
rs1800562 (C282Y)	*HFE*	A	G	A	Koller et al., 2016	European-American (USA)	Associated with TIBC	(34)
rs1800562 (C282Y)	*HFE*	A	G	G	Bedard et al., 2018	UK	G alleles associated with reduced iron stores	(55)
rs1800562 (C282Y)	*HFE*	A	G	A	McLaren et al., 2011	GWAS on Americans of European ancestry	Decreased TIBC and UIBC	(33)
rs1800562 (C282Y)	*HFE*	A	G	A	Beutler et al., 2003	European ancestry	Elevated Hb, TSAT, ferritin and lower anaemia prevalence	(14)

Table S2 continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs1800562 (C282Y)	*HFE*	A	G	A	Jackson et al., 2001	UK: Wales	Elevated serum ferritin, TSAT and Hb, and reduced UIBC; compound heterozygotes of H63D and C282Y has has high iron stores	(64)
rs1830084	*TF*	T	A	A	Benyamin, et al., 2009	Australian of European ancestry	increased transferrin	(18)
rs1830084	*TF*	T	A	T	Li et al., 2015	USA	Elevated TIBC	(12)
rs1867504	*TF*	A	G	A	Gichohi-Wainaina et al., 2016	Kenyans, Tanzanians, S. Africans and African-Americans	Elevated ferritin levels	(2)
rs1880669	*TF*	T	C	NA	McLaren et al., 2012	Whites, African-Americans, Hispanics and Asians (US & Canada)	Elevated TIBC in all the populations	(30)
rs1880669	*TF*	T	C	NA	Constantine et al., 2009	European ancestry	Reduced serum transferrin levels	(27)
rs1880669	*TF*	T	C	A	Piao et al., 2017	Chinese adolescents	A allele is associated with highet sTfR	(31)
rs198846	*HFE*	A	G	A	Kullo et al., 2010	USA	Elevated MCV and MCH	(22)
rs198846	*HFE*	A	G	A	Chambers et al., 2009	European and Indian Ancestry	The major allele (G) are associated with lower Hb concentration	(21)
rs2072860	*TMPRSS6*	G	A	NA	Bhathia et al., 2017	India	Associated with IRIDA	(43)
rs2072860	*TMPRSS6*	G	A	A	Li et al., 2015	USA	Elevated serum iron	(12)
rs2111833	*TMPRSS6*	T	C	T	McLaren et al., 2012	Whites, African-Americans, Hispanics and Asians (US & Canada)	Increased serum iron and TSAT in Whites	(30)

Table S2 Continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs2111833	*TMPRSS6*	T	C	T	Radio et al., 2016	Italy	No significant differences in iron biomarkers between genotypes	(4)
rs2160906	*TMPRSS6*	A	G	NA	Tanaka et al., 2010	Italy and USA	Significantly associated with reduced iron status	(57)
rs2235321	*TMPRSS6*	A	G	A	Lee et al., 2012	White Americans	Decreased TSAT	(38)
rs2235321	*TMPRSS6*	A	G	A	Delbini et al., 2010	Italy	Associated with iron deficiency	(39)
rs2235321	*TMPRSS6*	A	G	A	Poggiali et al., 2015	Italy	Associated with iron deficiecy	(40)
rs2235324	*TMPRSS6*	G	A	G	Lee et al., 2012	White Americans	Elevated TSAT	(38)
rs2235324	*TMPRSS6*	G	A	G	Beutler et al., 2010	Caucasians	Associated with iron deficiency	(51)
rs2235324	*TMPRSS6*	G	A	NA	Delbini et al., 2010	Italy	Associated with iron deficiency	(39)
rs2235324	*TMPRSS6*	G	A	NA	Tanaka et al., 2010	Italy and USA	Significantly associated with reduced iron status	(57)
rs2235324	*TMPRSS6*	G	A	NA	Poggiali et al., 2015	Italy	Associated with iron deficiecy	(40)
rs2280673	*TF*	A	C	NA	Benyamin, et al., 2009	Australian of European decent	Elevated transferrin, decrease TSAT and ferritin	(18)
rs228904	*TMPRSS6*	G	A	NA	Tanaka et al., 2010	Italy and USA	Significantly associated with elevated iron status	(57)
rs228916	*TMPRSS6*	C	T	T	Benyamin et al., 2014	European ancestry	Reduced serum iron	(5)
rs228918	*TMPRSS6*	A	G	C	Chambers et al., 2009	European and Indian Ancestry	Associated with decreased Hb levels, increased sTfR and low serum iron	(21)
rs228918	*TMPRSS6*	A	G	G	Gichohi-Wainaina et al., 2016	Kenyans, Tanzanians, S. Africans and African-Americans	Reduced Hb	(2)
rs228921	*TMPRSS6*	G	A	G	Chambers et al., 2009	European and Indian Ancestry	Associated with low iron status	(21)
rs228921	*TMPRSS6*	G	A	G	Gichohi-Wainaina et al., 2015	South Africa	Higher sTfR and lower serum iron levels in combination with rs228918	(35)

Table S2 Continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs2413450	*TMPRSS6*	T	C	T	Batar et al., 2018	Turkish	Associated with elevated TIBC	(47)
rs2413450	*TMPRSS6*	T	C	A	Gichohi-Wainaina et al., 2016	Kenyans, Tanzanians, S. Africans and African-Americans	Reduced Hb	(2)
rs2413450	*TMPRSS6*	T	C	T	Guo et al., 2016	Estonian	Elevated MCH	(63)
rs2413450	*TMPRSS6*	T	C	T	Ganesh et al, 2009	European ancestry	Reduced MCV, MCH and HCT	(53)
rs2543519	*TMPRSS6*	G	A	NA	Bhathia et al., 2017	India	Associated with IRIDA	(43)
rs2543519	*TMPRSS6*	G	A	NA	Delbini et al., 2010	Italians	Associated with iron deficiency	(39)
rs3811647	*TF*	A	G	A	Benyamin, et al., 2009	Australian of European dancestry	Increased transferrin	(18)
rs3811647	*TF*	A	G	A	Pichler et al., 2011	Italy and USA	Elevated transferrin concentration	(6)
rs3811647	*TF*	A	G	A	Blanco-Rojo et al., 2011	Spain	Elevated serum transferritin	(13)
rs3811647	*TF*	A	G	A	Constantine et al., 2009	European ancestry	Elevated serum transferritin	(27)
rs3811647	*TF*	A	G	A	An et al., 2012	Han Chinese	Reduced Hb, increased transferrin and TIBC	(32)
rs3811647	*TF*	A	G	A	Benyamin et al., 2009	Aurtralians	Raised serum iron and transferrin	(18)
rs3811647	*TF*	A	G	A	Gichohi-Wainaina et al., 2015	South Africa	Heterozygotes are associated with lower sTfR and higher iron stores; homozygotes of both extremes have similar levels of sTfR and body iron stores	(35)
rs3811647	*TF*	A	G	A	Li et al., 2015	USA	Elevated TIBC	(12)
rs3811647	*TF*	A	G	A	Koller et al., 2016	European-American (USA)	Associated with TIBC serum iron	(34)
rs3811647	*TF*	A	G	A	McLaren et al., 2011	Americans of European ancentry	Decreased TIBC and UIBC	(33)

Table S2 Continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs3811658	*TF*	T	C	NA	McLaren et al., 2012	Whites, African-Americans, Hispanics and Asians (US & Canada)	Elevated TIBC in all the populations	(30)
rs3811658	*TF*	T	C	T	Gichohi-Wainaina et al., 2016	Kenyans, Tanzanians, S. Africans and African-Americans	Increased Hb	(2)
rs3811658	*TF*	T	C	NA	Constantine et al., 2009	European ancestry	Increased transferrin	(27)
rs4820268	*TMPRSS6*	G	A	G	Benyamin, et al., 2009	Australian of European ancestry	Decreased serum iron and TSAT	(18)
rs4820268	*TMPRSS6*	G	A	G	Pichler et al., 2011	Italy and USA	Decreased serum iron, Hb, MCV, MCH and ferritin; increased TF, sTfR and sTfR/ferritin index	(6)
rs4820268	*TMPRSS6*	G	A	G	Kullo et al., 2010	USA	Associated with reduced MCH and MCHC	(22)
rs4820268	*TMPRSS6*	G	A	G	Constantine et al., 2009	European ancestry	Lower TSAT and serum iron	(27)
rs4820268	*TMPRSS6*	G	A	NA	Chambers et al., 2009	European and Indian Ancestry	Associated with decreased Hb levels, increased sTfR and low serum iron	(21)
rs4820268	*TMPRSS6*	G	A	NA	Delbini et al., 2010	Italians	Associated with iron deficiency	(39)
rs4820268	*TMPRSS6*	G	A	NA	Tanaka et al., 2010	Italy and USA	The C allele is significantly associated with increased serum iron and MCV, reduced RDW	(57)

Table S2 Continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs4820268	*TMPRSS6*	G	A	G	An et al., 2012	Han Chinese	Low Hb, serum iron, TSAT. Associated with the risk of IDA	(32)
rs4820268	*TMPRSS6*	G	A	NA	Ji et al., 2018	Australia	Associated with reduced ferritin levels	(62)
rs4820268	*TMPRSS6*	G	A	NA	Poggiali et al., 2015	Italy	Associated with iron deficiecy	(40)
rs4820268	*TMPRSS6*	G	A	NA	Gan et al., 2012	Chinese	Associated with low Hb and ferritin	(46)
rs4820268	*TMPRSS6*	G	A	A	Li et al., 2015	USA	Elevated serum iron and TSAT	(12)
rs4820268	*TMPRSS6*	G	A	G	Piao et al., 2017	Chinese adolescents	G alleles associated with lower serum ferritin	(31)
rs4820268	*TMPRSS6*	G	A	G	Gichohi-Wainaina et al., 2016	Kenyans, Tanzanians, S. Africans and African-Americans	Reduced Hb	(2)
rs5756504	*TMPRSS6*	T	C	T	Kullo et al., 2010	USA	Elevated MCH	(22)
rs5756504	*TMPRSS6*	T	C	NA	Tanaka et al., 2010	Italy and USA	The T allele Asscoiated with significant increase in serum iron levels	(57)
rs5756504	*TMPRSS6*	T	C	T	Seiki et al., 2018	Japanese	Associated with elevated MCV, MCH and MCHC	(60)
rs5756504	*TMPRSS6*	T	C	NA	Kamatani et al., 2010	Japanese	The T allele are asscoiated with elevated Hb	(56)
rs5756506	*TMPRSS6*	C	G	NA	Constantine et al., 2009	European ancestry	increase serum iron and TSAT	(27)
rs5756506	*TMPRSS6*	C	G	C	Seiki et al., 2018	Japanese	Elevated MCH, Hb	(60)
rs5756506	*TMPRSS6*	C	G	NA	Batar et al., 2018	Turkey	Associated with elevated Hb and HCT	(47)
rs5756512	*TMPRSS6*	T	C	NA	Bhathia et al., 2017	India	Associated with IRIDA	(43)
rs5756516	*TMPRSS6*	T	C	NA	Bhathia et al., 2017	India	Associated with IRIDA	(43)

Table S2 Continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs732756	*TMPRSS6*	C	T	NA	Bhathia et al., 2017	India	Associated with IRIDA	(43)
rs7385804	*TFR2*	C	A	C	Pichler et al., 2011	Italy and USA	Increased serum iron	(6)
rs7385804	*TFR2*	C	A	C	Benyamin et al., 2014	European ancestry	Elevated iron, TSAT and ferritin, and reduced tranferrin	(5)
rs7385804	*TFR2*	C	A	C	Soranzo et al., 2009	Europeans and South Asians	Elevated RBC	(37)
rs7385804	*TFR2*	C	A	C	An et al., 2012	Han Chinese	Lower TSAT and serum iron	(32)
rs7385804	*TFR2*	C	A	C	Gichohi-Wainaina et al., 2016	Kenyans, Tanzanians, S. Africans and African-Americans	Increased ferritin in Kenyans	(2)
rs744653	*SLC40A1*	C	T	T	Benyamin et al., 2014	European ancestry	Moderated HH via elevated transferrin and reduced TSAT, and ferritin levels	(5)
rs7638018	*TF*	G	A	NA	McLaren et al., 2012	Whites, African-Americans, Hispanics and Asians (US & Canada)	Increased TIBC in all the populations	(30)
rs78174698	*TMPRSS6*	A	G	NA	Bhathia et al., 2017	India	Associated with IRIDA	(43)
rs8177179	*TF*	G	A	A	Benyamin et al., 2014	European ancestry	Reduced transferrin	(5)
rs8177248	*TF*	T	C	NA	McLaren et al., 2012	Whites, African-Americans, Hispanics and Asians (US & Canada)	increased TIBC in all the populations	(30)
rs732756	*TMPRSS6*	C	T	NA	Bhathia et al., 2017	India	Associated with IRIDA	(12)

Table S2 Continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs855788	*TMPRSS6*	A	G	NA	Tanaka et al., 2010	Italy and USA	Significantly associated with elevated iron status	(57)
rs855791	*TMPRSS6*	A	G	A	Lee et al., 2012	White Americans	No significant assciation with outcome variables	(38)
rs855791	*TMPRSS6*	A	G	T	Pei et al., 2014	Taiwan	Associated with IDA; C alleles protective against IDA	(50)
rs855791	*TMPRSS6*	A	G	A	Athiyarath et al., 2015	India	Significantly associated with higher levels of serum iron	(10)
rs855791	*TMPRSS6*	A	G	A	Kullo et al., 2010	USA	Significantly associated with decreased MCV, MCH and MCHC	(22)
rs855791	*TMPRSS6*	A	G	A	Valenti et al., 2012	Italians	Lower MCV, ferritin and elevated hepcidin	(58)
rs855791	*TMPRSS6*	A	G	A	Beutler et al., 2010	European ancestry	Associated with iron deficiency	(51)
rs855791	*TMPRSS6*	A	G	A	Pelusi et al., 2013	Italy	Associated with elevated hepcidin; low MCV	(52)
rs855791	*TMPRSS6*	A	G	A	Chambers et al., 2009	European and Indian Ancestry	Associated with decreased Hb levels, increased sTfR and low serum iron	(21)
rs855791	*TMPRSS6*	A	G	NA	Bhathia et al., 2017	India	Associated with IRIDA	(43)
rs855791	*TMPRSS6*	A	G	NA	Delbini et al., 2010	Italy	Associated with iron deficiency	(39)
rs855791	*TMPRSS6*	A	G	NA	Tanaka et al., 2010	Italy and USA	The A allele is significantly associated with increased serum iron and Hb, MCV, reduced RDW	(57)
rs855791	*TMPRSS6*	A	G	NA	Galesloot et al., 2013	Netherlands	Significantly associated with serum iron and TSAT	(15)

Table S2 continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs855791	*TMPRSS6*	A	G	A	An et al., 2012	Han Chinese	Low Hb, serum iron, TSAT. Associated with the risk of IDA	(32)
rs855791	*TMPRSS6*	A	G	A	Benyamin et al., 2009	Aurtralians	Reduced serum iron, TSAT, ferritin, Hb and MCV, and raised transferrin	(18)
rs855791	*TMPRSS6*	A	G	A	van der Harst et al., 2012	Europeans and South Asians	Elevated MCH	(54)
rs855791	*TMPRSS6*	A	G	NA	Kamatani et al., 2010	Japanese	The G alleles are associated with elevated MCV, MCH and MCHC	(56)
rs855791	*TMPRSS6*	A	G	NA	Batar et al., 2018	Turkey	Associated with increased RBC count	(47)
rs855791	*TMPRSS6*	A	G	NA	Poggiali et al., 2015	Italy	Associated with low Hb, MCV and MCH	(40)
rs855791	*TMPRSS6*	A	G	NA	Gan et al., 2012	Chinese	Associated with low Hb and ferritin	(46)
rs855791	*TMPRSS6*	A	G	A	Nai et al., 2011	USA	Associated with elevated hepcidin, hepcidin/TSAT ratio, hepcidin/ferritin ratio, and low TSAT and serum iron	(48)
rs855791	*TMPRSS6*	A	G	A	Cheng et al., 2014	Australia	G alleles associated with higher serum iron and lower hepcidin at baseline	(59)
rs855791	*TMPRSS6*	A	G	NA	De Falco et al., 2018	Italy	Significanty associated with IDA	(16)
rs855791	*TMPRSS6*	A	G	T	Sorensen et al., 2015	Denmark	The T allele is associated with lower iron stores in men	(8)
rs855791	*TMPRSS6*	A	G	A	Benyamin et al., 2014	European ancestry	Reduced serum iron, TSAT and ferritin, and elevated tranferrin	(5)

Table S2 continued

SNPs	Gene	Minor allele	Major allele	Risk Allele	Study/ first author	Location/ Population/ Ethnicity[1]	Associated trait[2]	Reference[3]
rs855791	*TMPRSS6*	A	G	A	Pichler et al., 2013	European Ancestry populations	Protective against iron overload	(17)
rs855791	*TMPRSS6*	A	G	A	Traglia et al., 2011	Italians	Non-statistical significant increase in hepcidin/ferritin ratio and non-significant decrease in hepcidin and ferritin	(45)
rs855791	*TMPRSS6*	A	G	A	Danquah et al., 2014	Rwanda	Non-significantly associated with low Hb	(49)
rs855791	*TMPRSS6*	A	G	A	Bedard et al., 2018	UK	Associated with reduced iron stores	(55)
rs9610643	*TMPRSS6*	A	G	NA	Bhathia et al., 2017	India	Associated with IRIDA	(43)
rs9872999	*TF*	C	T	NA	Li et al., 2015	USA	Reduced TIBC	(12)

Hb, haemoglobin; HCT, haematocrit; HFE, High fe; HH, hereditary hemochromatosis; IDA, iron deficiency anaemia; MCH, mean corpuscular haemoglobin; NA, not available; SAS, South Asians; SI, serum iron; *SLC40A1*, solute carrier family 40 member 1; SNP, single nucleotide polymorphism; sTfR, soluble transferrin receptor
NA indicates SNPs that we could not establish the risk allele because it was not stated by the respective studies that reported the SNPs.

[1]The population or the study location where the study that reported each SNP was conducted.
[2]The phenotype that was reported
[3]The study that reported each SNP

Reference:

1. Andreani M, Radio FC, Testi M, De Bernardo C, Troiano M, Majore S, et al. Association of hepcidin promoter c.-582 A>G variant and iron overload in thalassemia major. Haematologica. 2009;94(9):1293–6.
2. Gichohi-Wainaina WN, Tanaka T, Towers GW, Verhoef H, Veenemans J, Talsma EF, et al. Associations between Common Variants in Iron-Related Genes with Haematological Traits in Populations of African Ancestry. PLoS One [Internet]. 2016;11(6):e0157996. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27332551
3. Javaheri-Kermani M, Farazmandfar T, Ajami A, Yazdani Y. Impact of hepcidin antimicrobial peptide on iron overload in tuberculosis patients. Scand J Infect Dis [Internet]. 2014;46(10):693–6. Available from: http://informahealthcare.com/doi/abs/10.3109/00365548.2014.929736
4. Radio FC, Majore S, Aurizi C, Sorge F, Biolcati G, Bernabini S, et al. Hereditary hemochromatosis type 1 phenotype modifiers in Italian patients. The controversial role of variants in HAMP, BMP2, FTL and SLC40A1 genes. Blood Cells Mol Dis [Internet]. 2015 Jun;55(1):71–5. Available from: http://dx.doi.org/10.1016/j.bcmd.2015.04.001
5. Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. Nat Commun [Internet]. 2014 Oct 29;5(2):4926. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25352340
6. Pichler I, Minelli C, Sanna S, Tanaka T, Schwienbacher C, Naitza S, et al. Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels. Hum Mol Genet [Internet]. 2011

Mar 15;20(6):1232–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21208937

7. Garewal G, Das R, Ahluwalia J, Marwaha RK. Prevalence of the H63D mutation of the HFE in north India: Its presence does not cause iron overload in beta thalassemia trait. Eur J Haematol. 2005;74(4):333–6.
8. Sørensen E, Rigas AS, Thørner LW, Burgdorf KS, Pedersen OB, Petersen MS, et al. Genetic factors influencing ferritin levels in 14,126 blood donors: Results from the Danish Blood Donor Study. Transfusion. 2016;56(3):622–7.
9. Mast AE, Lee T-H, Schlumpf KS, Wright DJ, Johnson B, Carrick DM, et al. The impact of HFE mutations on haemoglobin and iron status in individuals experiencing repeated iron loss through blood donation*. Br J Haematol [Internet]. 2012 Feb;156(3):388–401. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22118647
10. Athiyarath R, Shaktivel K, Abraham V, Singh D, Bondu JD, Chapla A, et al. Association of genetic variants with response to iron supplements in pregnancy. Genes Nutr [Internet]. 2015 Jul 30;10(4):25. Available from: http://link.springer.com/10.1007/s12263-015-0474-2
11. Whitfield JB, Cullen LM, Jazwinska EC, Powell LW, Heath AC, Zhu G, et al. Effects of HFE C282Y and H63D polymorphisms and polygenic background on iron stores in a large community sample of twins. Am J Hum Genet [Internet]. 2000 Apr;66(4):1246–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10739755
12. Li J, Lange LA, Duan Q, Lu Y, Singleton AB, Zonderman AB, et al. Genome-wide admixture and association study of serum iron, ferritin, transferrin saturation and total iron binding capacity in African Americans. Hum Mol

Genet. 2015;24(2):572–81.

13. Blanco-Rojo R, Baeza-Richer C, López-Parra AM, Pérez-Granados AM, Brichs A, Bertoncini S, et al. Four variants in transferrin and HFE genes as potential markers of iron deficiency anaemia risk: an association study in menstruating women. Nutr Metab (Lond) [Internet]. 2011 Oct 6;8:69. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21978626
14. Beutler E, Felitti V, Gelbart T, Waalen J. Haematological effects of the C282Y HFE mutation in homozygous and heterozygous states among subjects of northern and southern European ancestry. Br J Haematol. 2003;120(5):887–93.
15. Galesloot TE, Geurts-Moespot AJ, den Heijer M, Sweep FCGJ, Fleming RE, Kiemeney L a LM, et al. Associations of common variants in HFE and TMPRSS6 with iron parameters are independent of serum hepcidin in a general population: a replication study. J Med Genet [Internet]. 2013;50(9):593–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23794717
16. De Falco L, Tortora R, Imperatore N, Bruno M, Capasso M, Girelli D, et al. The role of TMPRSS6 and HFE variants in iron deficiency anemia in celiac disease. Am J Hematol. 2018;93(3):383–93.
17. Pichler I, Del Greco M F, Gögele M, Lill CM, Bertram L, Do CB, et al. Serum iron levels and the risk of Parkinson disease: a Mendelian randomization study. PLoS Med [Internet]. 2013;10(6):e1001462. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23750121
18. Benyamin B, McRae AF, Zhu G, Gordon S, Henders AK, Palotie A, et al. Variants in TF and HFE explain approximately 40% of genetic variation in

serum-transferrin levels. Am J Hum Genet [Internet]. 2009 Jan;84(1):60–5. Available from: http://dx.doi.org/10.1016/j.ajhg.2008.11.011

19. Gordeuk VR, Brannon PM. Ethnic and genetic factors of iron status in women of reproductive age. Am J Clin Nutr. 2017;106:1594S-1599S.
20. Blanco-Rojo R, Toxqui L, López-Parra AM, Baeza-Richer C, Pérez-Granados AM, Arroyo-Pardo E, et al. Influence of diet, menstruation and genetic factors on iron status: A cross-sectional study in Spanish women of childbearing age. Int J Mol Sci. 2014;15(3):4077–87.
21. Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, et al. Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. Nat Genet [Internet]. 2009 Nov;41(11):1170–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19820698
22. Kullo IJ, Ding K, Jouni H, Smith CY, Chute CG. A Genome-Wide Association Study of Red Blood Cell Traits Using the Electronic Medical Record. 2010;5(9):1–9.
23. Chen Z, Tang H, Qayyum R, Schick UM, Nalls MA, Handsaker R, et al. Genome-wide association analysis of red blood cell traits in African Americans: The cogent network. Hum Mol Genet. 2013;22(12):2529–38.
24. Kasvosve I, Gomo ZAR, Nathoo KJ, Matibe P, Mudenge B, Loyevsky M, et al. Effect of ferroportin Q248H polymorphism on iron status in African children. 2018;(April):1102–6.
25. Masaisa F, Breman C, Gahutu JB, Mukiibi J, Delanghe J, Philippé J. Ferroportin (SLC40A1) Q248H mutation is associated with lower circulating serum hepcidin levels in Rwandese HIV-positive women. Ann Hematol. 2012;91(6):911–6.

26. Rivers CA, Barton JC, Gordeuk VR, Acton RT, Speechley MR, Snively BM, et al. Association of ferroportin Q248H polymorphism with elevated levels of serum ferritin in African Americans in the Hemochromatosis and Iron Overload Screening (HEIRS) Study. Blood Cells, Mol Dis. 2007;38(3):247–52.
27. Constantine CC, Anderson GJ, Vulpe CD, Mclaren CE, Bahlo M, Yeap HL, et al. A novel association between a SNP in CYBRD1 and serum ferritin levels in a cohort study of HFE hereditary haemochromatosis. 2009;(August):140–9.
28. Sarria B, Lopez-parra AM, Perez-granados AM, Arroyo-pardo E, Roe MA, Teucher B, et al. The G277S transferrin mutation does not affect iron absorption in iron deficient women. 2007;57–60.
29. Lee PL, Halloran C, Trevino R, Felitti V, Beutler E. Human transferrin G277S mutation: a risk factor for iron deficiency anaemia. Br J Haematol [Internet]. 2001 Nov;115(2):329–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11703331
30. McLaren CE, McLachlan S, Garner CP, Vulpe CD, Gordeuk VR, Eckfeldt JH, et al. Associations between single nucleotide polymorphisms in iron-related genes and iron status in multiethnic populations. PLoS One. 2012;7(6).
31. Piao W, Wang L, Zhang T, Wang Z, Shangguan S, Sun J, et al. A single-nucleotide polymorphism in transferrin is associated with soluble transferrin receptor in Chinese adolescents. Asia Pac J Clin Nutr. 2017;26(6):1170–8.
32. An P, Wu Q, Wang H, Guan Y, Mu M, Liao Y, et al. TMPRSS6, but not TF, TFR2 or BMP2 variants are associated with increased risk of iron-deficiency anemia. Hum Mol Genet. 2012;21(9):2124–31.
33. McLaren CE, Garner CP, Constantine CC, McLachlan S, Vulpe CD, Snively BM, et al. Genome-wide association study identifies genetic loci associated

with iron deficiency. PLoS One. 2011;6(3).

34. Koller DL, Imel EA, Lai D, Padgett LR, Acton D, Gray A, et al. Genome-wide association study of serum iron phenotypes in premenopausal women of European descent. Blood Cells, Mol Dis [Internet]. 2016;57:50–3. Available from: http://dx.doi.org/10.1016/j.bcmd.2015.12.002

35. Gichohi-Wainaina, W. N. Melse-Boonstra, A. Swinkels, D. W. Zimmermann, M. B. Feskens, E. J. Towers GW. Common variants and haplotypes in the TF, TNF- alpha , and TMPRSS6 genes are associated with iron status in a female black South. J Nutr 2015. 2015;145(5):945–53.

36. Benyamin B, Ferreira MAR, Willemsen G, Gordon S, Middelberg RPS, McEvoy BP, et al. Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. Nat Genet [Internet]. 2009 Nov 11;41(11):1173–5. Available from: http://www.nature.com/doifinder/10.1038/ng.456

37. Soranzo N, Spector TD, Mangino M, Kühnel B, Rendon A, Teumer A, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet [Internet]. 2009;41(11):1182–90. Available from: http://dx.doi.org/10.1038/ng.467

38. Lee PL, Barton JC, Khaw PL, Bhattacharjee SY, Barton JC. Common TMPRSS6 mutations and iron, erythrocyte, and pica phenotypes in 48 women with iron deficiency or depletion. Blood Cells, Mol Dis [Internet]. 2012 Feb;48(2):124–7. Available from: http://dx.doi.org/10.1016/j.bcmd.2011.12.003

39. Delbini P, Vaja V, Graziadei G, Duca L, Nava I, Refaldi C, et al. Genetic variability of TMPRSS6 and its association with iron deficiency anaemia. Br J Haematol. 2010;151(3):281–4.

40. Poggiali E, Andreozzi F, Nava I, Consonni D, Graziadei G, Cappellini MD. The role of TMPRSS6 polymorphisms in iron deficiency anemia partially responsive to oral iron treatment. Am J Hematol. 2015;90(4):306–9.

41. Galesloot TE, Verweij N, Traglia M, Barbieri C, Van Dijk F, Geurts-Moespot AJ, et al. Meta-GWAS and meta-analysis of exome array studies do not reveal genetic determinants of serum hepcidin. PLoS One. 2016;11(11):1–13.

42. Kloss-Brandstätter A, Erhart G, Lamina C, Meister B, Haun M, Coassin S, et al. Candidate gene sequencing of SLC11A2 and TMPRSS6 in a family with severe anaemia: Common SNPs, rare haplotypes, no causative mutation. PLoS One. 2012;7(4):1–8.

43. Bhatia P, Singh A, Hegde A, Jain R, Bansal D. Systematic evaluation of paediatric cohort with iron refractory iron deficiency anaemia (IRIDA) phenotype reveals multiple TMPRSS6 gene variations. Br J Haematol [Internet]. 2017 Apr;177(2):311–8. Available from: http://doi.wiley.com/10.1111/bjh.14554

44. Valenti L, Fracanzani AL, Rametta R, Fraquelli M, Soverini G, Pelusi S, et al. Effect of the A736V TMPRSS6 polymorphism on the penetrance and clinical expression of hereditary hemochromatosis. J Hepatol [Internet]. 2012;57(6):1319–25. Available from: http://dx.doi.org/10.1016/j.jhep.2012.07.041

45. Traglia M, Girelli D, Biino G, Campostrini N, Corbella M, Sala C, et al. Association of HFE and TMPRSS6 genetic variants with iron and erythrocyte parameters is only in part dependent on serum hepcidin concentrations. J Med Genet [Internet]. 2011 Sep;48(9):629–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21785125

46. Gan W, Guan Y, Wu Q, An P, Zhu J, Lu L, et al. Association of TMPRSS6 polymorphisms with ferritin, hemoglobin, and type 2 diabetes risk in a Chinese Han population. Am J Clin Nutr [Internet]. 2012 Mar;95(3):626–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22301935

47. Batar B, Bavunoglu I, Hacioglu Y, Cengiz M, Mutlu T, Yavuzer S, et al. The role of TMPRSS6 gene variants in iron-related hematological parameters in Turkish patients with iron deficiency anemia. Gene [Internet]. 2018;673(January):201–5. Available from: https://doi.org/10.1016/j.gene.2018.06.055

48. Nai A, Pagani A, Silvestri L, Campostrini N, Corbella M, Girelli D, et al. TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. Blood [Internet]. 2011 Oct 20;118(16):4459–62. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2011-06-364034

49. Danquah I, Gahutu J-B, Zeile I, Musemakweri A, Mockenhaupt FP. Anaemia, iron deficiency and a common polymorphism of iron-regulation, TMPRSS6 rs855791, in Rwandan children. Trop Med Int Health [Internet]. 2014;19(1):117–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24175968

50. Pei SN, Ma MC, You HL, Fu HC, Kuo CY, Rau KM, et al. TMPRSS6 rs855791 polymorphism influences the susceptibility to iron deficiency anemia in women at reproductive age. Int J Med Sci. 2014;11(6):614–9.

51. Beutler E, Van Geet C, te Loo DMWM, Gelbart T, Crain K, Truksa J, et al. Polymorphisms and mutations of human TMPRSS6 in iron deficiency anemia. Blood Cells, Mol Dis [Internet]. 2010 Jan 15;44(1):16–21. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/19818657

52. Pelusi S, Girelli D, Rametta R, Campostrini N, Alfieri C, Traglia M, et al. The A736V TMPRSS6 polymorphism influences hepcidin and iron metabolism in chronic hemodialysis patients: TMPRSS6 and hepcidin in hemodialysis. BMC Nephrol [Internet]. 2013;14:48. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3585892&tool=pmcentrez&rendertype=abstract

53. Ganesh SK, Zakai NA, van Rooij FJA, Soranzo N, Smith A V, Nalls MA, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet [Internet]. 2009 Nov 11;41(11):1191–8. Available from: http://dx.doi.org/10.1038/ng.466

54. van der Harst P, Zhang W, Mateo Leach I, Rendon A, Verweij N, Sehmi J, et al. Seventy-five genetic loci influencing the human red blood cell. Nature. 2013;492(7429):369–75.

55. Bédard A, Lewis SJ, Burgess S, John Henderson A, Shaheen SO. Maternal iron status during pregnancy and respiratory and atopic outcomes in the offspring: A Mendelian randomisation study. BMJ Open Respir Res. 2018;5(1):1–10.

56. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet [Internet]. 2010;42(3):210–5. Available from: http://dx.doi.org/10.1038/ng.531

57. Tanaka T, Roy CN, Yao W, Matteini A, Semba RD, Arking D, et al. A genome-wide association analysis of serum iron concentrations. Blood [Internet]. 2010 Jan 7;115(1):94–6. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/19880490

58. Valenti L, Rametta R, Dongiovanni P, Motta BM, Canavesi E, Pelusi S, et al. The A736V TMPRSS6 Polymorphism Influences Hepatic Iron Overload in Nonalcoholic Fatty Liver Disease. PLoS One. 2012;7(11).

59. Cheng HL, Hancock DP, Rooney KB, Steinbeck KS, Grif HJ, Connor HTO. SHORT COMMUNICATION A candidate gene approach for identifying differential iron responses in young overweight women to an energy-restricted haem iron-rich diet. 2014;(February):1250–2.

60. Seiki T, Naito M, Hishida A, Takagi S, Matsunaga T, Sasakabe T, et al. Association of genetic polymorphisms with erythrocyte traits: Verification of SNPs reported in a previous GWAS in a Japanese population. Gene [Internet]. 2018;642(October 2017):172–7. Available from: http://dx.doi.org/10.1016/j.gene.2017.11.031

61. Alfred T, Ben-Shlomo Y, Cooper R, Hardy R, Deary IJ, Elliott J, et al. Genetic variants influencing biomarkers of nutrition are not associated with cognitive capability in middle-aged and older adults. J Nutr [Internet]. 2013 May;143(5):606–12. Available from: http://jn.nutrition.org/cgi/doi/10.3945/jn.112.171520

62. Ji Y, Flower R, Hyland C, Saiepour N, Faddy H. Genetic factors associated with iron storage in Australian blood donors. Blood Transfus. 2018;16(2):123–9.

63. Guo MH, Nandakumar SK, Ulirsch JC, Zekavat SM, Buenrostro JD, Natarajan P, et al. Comprehensive population-based genome sequencing provides insight into hematopoietic regulatory mechanisms. Proc Natl Acad Sci [Internet]. 2017;114(3):E327–36. Available from:

http://www.pnas.org/lookup/doi/10.1073/pnas.1619052114

64. Jackson HA, Carter K, Darke C, Guttridge MG, Ravine D, Hutton RD, et al. HFE mutations, iron deficiency and overload in 10 500 blood donors. Br J Haematol. 2001;114(2):474–84.

Table S3: Population Branch Statistic (PBS) values involving the comparison of three populations.

SNP	PBS-AFR	PBS-EUR	PBS-SAS
rs10421768	0.001	0.003	0.000
rs744653	0.001	0.003	0.000
rs11568350	0.021	0.000	0.000
rs1439816	0.266	0.051	0.000
rs2235321	0.000	0.003	0.027
rs855791	0.163	0.000	0.071
rs78174698	0.000	0.011	0.051
rs5756504	0.109	0.000	0.060
rs5756506	0.249	0.000	0.078
rs4820268	0.045	0.000	0.043
rs2413450	0.116	0.000	0.057
rs2072860	0.098	0.000	0.035
rs9610643	0.098	0.000	0.035
rs855788	0.286	0.000	0.042
rs2543519	0.031	0.009	0.000
rs2111833	0.000	0.002	0.025
rs2235324	0.000	0.001	0.002
rs1421312	0.017	0.013	0.000
rs5756512	0.011	0.010	0.000
rs2160906	0.026	0.013	0.000
rs732756	0.017	0.012	0.000
rs228904	0.016	0.012	0.000
rs11704654	0.001	0.002	0.000
rs5756516	0.000	0.030	0.027
rs228916	0.027	0.006	0.000
rs228918	0.001	0.000	0.006
rs228921	0.001	0.000	0.006
rs1867504	0.089	0.000	0.009
rs9872999	0.000	0.029	0.026
rs8177179	0.000	0.024	0.022
rs12493168	0.047	0.019	0.000
rs1799852	0.034	0.000	0.017
rs1799899	0.033	0.005	0.000
rs3811658	0.093	0.000	0.021
rs8177248	0.130	0.000	0.024
rs8177253	0.032	0.000	0.014
rs1405023	0.040	0.000	0.013
rs1880669	0.054	0.013	0.000

rs3811647	0.042	0.000	0.014
rs1358024	0.125	0.000	0.033
rs1525892	0.014	0.000	0.010
rs1049296	0.032	0.000	0.028
rs7638018	0.071	0.000	0.019
rs1830084	0.093	0.000	0.019
rs2280673	0.000	0.010	0.027
rs1799945	0.043	0.043	0.000
rs1800562	0.001	0.020	0.000
rs198846	0.000	0.013	0.012
rs129128	0.052	0.026	0.000
rs7385804	0.000	0.003	0.000
Average	0.053	0.008	0.017

(AFR=Africa, EUR- Europe, SAS=South Asia)

Chapter 4:

Association between common *TMPRSS6* and *TF* gene variants with hepcidin and iron status in healthy rural Gambians

Chapter description:

This chapter presents the results of the cross-sectional study of the effects of common *TMPRSS6* and *TF* SNPs on hepcidin and iron status in healthy Gambians.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1513421	**Title**	Mr.
First Name(s)	Momodou W.		
Surname/Family Name	Jallow		
Thesis Title	The impact of single nucleotide polymorphisms in human genes that regulate hepcidin and iron on oral iron absorption and the risk of anaemia in Africans		
Primary Supervisor	Dr Susana Campino		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	**N/A**		
When was the work published?	**N/A**		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be Published?	**Scientific Reports**
Please list the paper's authors in the intended authorship order:	**Momodou W. Jallow, Susana Campino, Andrew M Prentice and Carla Cerami**

Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	**I am the main corresponding author. I contributed to conceptualisation and design of this study. I did the literature review and data analysis. I drafted the manuscript, managed co-author comments, and the submission process.**

SECTION E

Student Signature	[redacted]
Date	**15 January 2021**

Supervisor Signature	[redacted]
Date	**15 January 2021**

Title

Association of common *TMPRSS6* and *TF* gene variants with hepcidin and iron status in healthy rural Gambians

Authors and Affiliations

Momodou W. Jallow,[1,2] Susana Campino,[2] Andrew M. Prentice[1] and Carla Cerami,[1*]

[1]Nutrition Theme, MRC Unit The Gambia at London School of Hygiene & Tropical Medicine, Atlantic Boulevard, Fajara, P.O. Box 273, Banjul, The Gambia; [2]Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

*Address correspondence to CC (Carla.cerami@mrc.gm).

4.1. Abstract

Background: Genome-wide association studies in Europeans and Asians have identified numerous variants in the transmembrane protease serine 6 (*TMPRSS6*) and transferrin (*TF*) genes that are associated with changes in iron status.

Aims: We sought to investigate the effects of common *TMPRSS6* and *TF* gene SNPs on iron status indicators in a cohort of healthy Africans from rural Gambia.

Methods: We measured iron biomarkers and haematology traits on individuals participating in the Keneba Biobank with genotype data on *TMPRSS6* (rs2235321, rs855791, rs4820268, rs2235324, rs2413450 and rs5756506) and *TF* (rs3811647 and rs1799852), n=1316. After controlling for inflammation, age and sex, we analysed the effects of carrying either single or multiple iron-lowering alleles on iron status.

Results: *TMPRSS6* rs2235321 was significantly associated with plasma hepcidin concentrations (AA genotypes having lower hepcidin levels; F ratio 3.7, P=0.014) with greater impact in individuals with low haemoglobin or ferritin. No other *TMPRSS6* variant affected influenced plasma hepcidin levels. None of the *TMPRSS6* variants nor a *TMPRSS6* allele risk score affected other iron biomarkers or haematological traits. *TF* rs3811647 AA carriers had 21% higher transferrin (F ratio 16.0, P<0.0001), 24% higher unsaturated iron-binding capacity (F ratio 12.8, P<0.0001) and 25% lower transferrin saturation (TSAT) (F ratio 4.3, P<0.0001) compared to GG carriers. There was no association between either *TF* SNP and any haematological traits or iron biomarkers.

Conclusions: We identified an association between *TMPRSS6* rs2235321 and plasma hepcidin levels and replicated the previous findings on the effects of *TF* rs3811647 on transferrin and iron binding capacity. However, the effects are subtle and contribute little to population variance. Further genetic and functional studies,

including polymorphisms frequent in Africa populations, are needed to identify markers for genetically stratified approaches to prevention or treatment of iron deficiency anaemia.

Keywords: Genetic variations; *TMPRSS6*, *TF*, iron status, risk alleles

4.2. Introduction

The discovery of hepcidin and the molecular mechanisms modulating its function in iron metabolism have brought new insights into how iron is regulated in the human body [1,2]. Subsequently, several genome-wide studies (GWASs) have revealed single nucleotide polymorphism (SNPs) in genes involved in hepcidin regulatory pathways, that are associated with impaired iron status [3–5]. The most common SNPs associated with low iron status are in the *TMPRSS6* gene, encoding the matriptase-2 protein [6–8]. *TMPRSS6* suppresses hepcidin synthesis, and its impaired function has been associated with inappropriately high hepcidin, which restricts iron absorption by the duodenum and iron mobilisation from storage [6,9,10]. Impaired *TMPRSS6* activity has been implicated in the development of iron-refractory iron deficiency anaemia (IRIDA) [8].

So far, more than 50 SNPs within the *TMPRSS6* gene have been reported to be associated with impaired iron status. The most commonly reported SNPs are rs855791 and rs4820268 and rs2235321 [3,4,11–14]. However, most studies linking *TMPRSS6* SNPs and low iron status were conducted in Europeans and Asians. Genetic variations in the *TMPRSS6* gene has been linked to variations in iron status indicators in different populations across the world, including in India [12], Turkey [15] and Australia [16].

Also, SNPs in the transferrin (*TF*) gene have been associated with altered iron status [17–19]. The most common *TF* SNP associated with the risk of iron deficiency is rs3811647 [17,20–22]. This SNP is associated with low iron status in different populations globally, including in African populations [23]. However, little information exists on the effects of *TF* SNPs on low iron status, particularly in settings with high anaemia burden.

Despite efforts to identify genetic risk factors for anaemia, very few such studies have been reported from sub-Saharan Africa. Exceptions are one study on the effects of *TMPRSS6, TF*, and *tumour necrosis factor (TNF)- alpha* on iron status in Black South African Women [23] and a similar study on the effects of TF and TMPRSS6 SNPs on haematological phenotypes in four African ancestry cohorts (Kenyans, Tanzanians, South Africans and African Americans) [24]. To the best of our knowledge, no study has been done in West Africa to assess the effects of genetic variants in hepcidin and iron regulatory genes on low iron status. This is particularly important given that West Africa is one of the regions with the highest prevalence of anaemia [25]. In this study, we investigated the association between common SNPs in the *TMPRSS6* and *TF* genes, and iron indicators in healthy individuals from the rural Gambia.

4.3. Subjects and Methods

Study populations and sample collection

This study utilised a cohort of healthy individuals, enrolled in the Keneba Biobank at MRCG@LSHTM [26]. Based on the availability of genotype data, we studied 1316 individuals aged 1 to 87 years (54.2% females). Each participant was interviewed and had a basic health examination, and those with significant health conditions excluded. After an overnight fast, a venous blood sample was collected in EDTA and lithium heparin tubes. DNA was extracted from cell pellets using standard procedures and stored at -70^{o}C. Plasma samples from lithium heparin anticoagulant were stored at -70^{o}C freezers until analysis.

Haematology measurements

Full blood count (FBC) was performed within 4 hours of sample collection using a Medonic M-Series automated haematology analyser (Boule Medical, Sweden) and results analysed for haemoglobin (Hb), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), mean corpuscular Hb concentration (MCHC), red cell distribution width (RDW), red blood cell number (RBC) and haematocrit (Hct).

Biomarker measurements

Iron biomarkers [serum iron, transferrin, ferritin, unsaturated iron-binding capacity (UIBC), soluble transferrin receptor (sTfR) and the inflammation marker C-reactive protein (CRP)] were measured using an automated biochemistry analyser (COBAS Integra 400 Plus; Roche Diagnostics). For all the biochemistry analysis, a single 500µl plasma aliquot was used to measure all the iron parameters. The analyser was calibrated using commercial calibrators for each parameter. Also, to ensure quality and reliable results, commercially available standards were analysed alongside the samples. Haemolysed samples are excluded from the analysis. Plasma total iron-binding capacity (TIBC) and transferrin saturation (TSAT) were calculated from UIBC and plasma iron (TIBC = plasma iron + UIBC) and TSAT= [plasma iron/TIBC] X 100).

Hepcidin quantification

Plasma hepcidin was quantified using a competitive ELISA (Bachem Hepcidin-25 [human], EIA kit; Peninsula Laboratories International), with a detection range of 0.049–25 ng/mL, as previously described [27]. Hepcidin was measured in duplicates and the mean of the two reads was calculated. Samples with a coefficient of variation (CV) greater than 10% were repeated. Concentrations were read at 450nm wavelength

using a MultiskanTM FC Microplate Photometer (Thermofisher Scientific). Concentrations were interpolated from a 4-parameter curve fitted from a 2-fold, 10-point serial dilution made from a manufacturer-provided standard peptide. The concentrations were calculated using the SkanIt Software (Thermofisher Scientific). Samples outside the standard curve were re-analyzed at a higher dilution, and the final concentration was calculated based on the dilution factor.

Genotyping

This study population was genotyped using the Illumina Infinium 240K Human Exome Beadchip (v1.0 and v1.1), as previously described [28], in which 848 SNPs were genotyped. Genotype calling was performed using data-driven clustering (Genome Studio, Illumina, CA, USA). The *TMPRSS6* rs2235321, rs855791, rs4820268, rs2235324, rs2413450 and rs5756506, and *TF* rs3811647 and rs1799852 were selected for inclusion in this study based on their previously published association with iron status.

Genotype combinations and allele risk scores

For both the *TMPRSS6* and *TF* SNPs, we generated genotype combinations and allele risk scores (ARS) by summing up the genotypes and the number of risk alleles respectively, from all the SNPs an individual carried. Risk alleles were defined as the alleles that are previously reported to be associated with low iron at each SNP. For each SNP, genotypes were assigned 0,1 or 2, with risk alleles assigned 1 and the alternate allele assigned 0. Thus, homozygous for the risk allele scored 2 and homozygous for non-risk alleles were scored 0, **Table S1** (*TMPRSS6* SNPs) and **Table S2** (*TF* SNPs).

A total of 94 genotype combinations from the six *TMPRSS6* SNPs were found in our population (**Table S3**)**.** We investigated the effects of individual *TMPRSS6* and *TF* SNPs on all the iron biomarkers and haematology phenotypes. In addition, based on the functional role of *TMPRSS6* on hepcidin regulation, and its effects on iron status, we investigated the effects of *TMPRSS6* ARS on hepcidin. Similarly, we assessed the effects of *TF* SNPs' genotype combinations and *TF* ARS on transferrin level.

Statistical analysis

The effects of genetic variants (genotypes of single SNPs or combinations of multiple SNPs) on iron biomarkers were determined by linear modelling with iron and haematological traits as response variables and genotype as dependent variables. Age, sex, inflammation (CRP) were added as covariates where indicated. We tested the effects each SNP individually and in combinations on iron biomarkers. Hepcidin, ferritin and CRP were log transformed. We added log ferritin as a covariate when analysing the effects of genotype on plasma hepcidin because *TMPRSS6* modulates the interaction between iron status and hepcidin gene expression. Furthermore, we stratified the study population based on haemoglobin and ferritin levels and determined the effects of genotype on each sub-population. For each sub-population, we used analysis of variance (ANOVA) to determine the effects of individual SNPs on iron biomarkers. Bonferroni correction was applied to account for multiple testing. The statistical analyses were conducted using R statistical software [29] and DataDesk Version 7.0.2 (Data Description Inc, Ithaca).

Ethics

The study was approved by the MRCG@LSHTM / Gambia Government Ethics Committee (SCC1185). A written informed consent was obtained from each study participant.

4.4. Results

Baseline characteristics of the study population are presented in **Table 1**. Due mostly to out-migration of males there was a slight sex bias (54.2% were female). There were significant differences between the sexes in age, RBC count and RBC indices, serum iron and transferrin.

Table 2. Demographic characteristics of the study population

Variables	All (n=1316)	Males (n=595)	Female (n=721)	*P*-value
Age, median (range)	9 (1, 87)	11.5 (1, 79)	19.7 (1, 87)	0.000
Gender, M/F (%)	45.8/54.2	-	-	-
Hb (g/dl)	11.6 (6.5, 16.0)	11.6 (8.2, 16.0)	11.6 (6.5, 15.3)	0.056
RBC (x10^12)	4.20 (2.41, 5.62)	4.30 (2.90, 5.61)	4.21 (2.41, 5.65)	0.000
MCV (fl)	78.9 (51.3, 103.2)	78.2 (51.3, 103.2)	79.6 (52.7, 97.9)	0.000
Haematocrit (%)	33.0 (20.1, 48.10)	32.9 (24.3, 48.1)	33.1 (20.1, 44.9)	0.374
RDW (%)	14.6 (12.7, 27.4)	14.6 (12.8, 24.0)	14.5 (12.7, 27.4)	0.320
MCH (pg)	27.6 (16.2, 34.9)	27.4 (16.2, 34.2)	27.9 (17.3, 34.9)	0.005
MCHC (g/dl)	34.9 (28.9, 37.2)	35.1 (29.9, 37.2)	34.8 (28.9, 37.1)	0.024
Serum iron (umol/l)	12.15 (0.60, 52.40)	11.85 (1.7, 28.9)	12.50 (0.6, 52.4)	0.005
Hepcidin (ng/ml)	8.86 (0.11, 103.78)	8.70 (0.17, 94.64)	9.00 (0.11, 103.78)	0.780
TSAT (%)	20.73 (2.84, 75.72)	20.07 (2.84, 57.51)	21.42 (4.31, 75.72)	0.114
Transferrin (g/l)	2.75 (0.01, 4.77)	2.74 (0.00, 4.21)	2.77 (0.00, 4.77)	0.009
TIBC (umol/l)	60.4 (1.4, 129.3)	60.0 (20.5, 129.3)	60.8 (1.4, 123.0)	0.028
UIBC (umol/l)	47.3 (0.8, 120.8)	47.3 (11.8, 120.8)	47.2 (0.8, 113.4)	0.642
Ferritin (ug/l)	26.9 (0.10 (166.8)	27.7 (0.2, 161.7)	25.3 (0.1, 166.8)	0.063
sTfR (mg/L)	4.93 (0.70, 19.77)	4.84 (0.79, 14.11)	4.41 (0.00, 19.77)	0.083

CRP (mg/l)	1.19 (0.0, 40.26)	1.11 (0.00, 40.26)	1.23 (0.00, 32.24)	0.083

Data are presented in median (ranges), except gender.
CRP, C-reactive protein; Hb, haemoglobin; MCV. Mean corpuscular volume; MCH, mean corpuscular haemoglobin; MCHC, mean corpuscular Hb concentration; RDW, red cell distribution width; RBC, red blood cells; ; sTfR, soluble transferrin receptor; TIBC, total iron-binding capacity; UIBC unsaturated iron-binding capacity; TSAT, transferrin saturation.

TMPRSS6 variants

All the SNPs investigated were in Hardy-Weinberg Equilibrium. Also, all were in low linkage disequilibrium (LD) in this study population, except rs4820268 and rs2413450 which have r^2=0.7 (**Figure 1**)**.** Among the SNPs we investigated, *TMPRSS6* rs2235324 had the highest minor allele frequency (MAF) in our study population (45%), and *TMPRSS6* rs855791 and *TF* rs1799852 had the lowest MAF (7% each) [28]. There was no detectable influence of sex on any associations, so sex was discarded from the models. None of the *TMPRSS6* SNPs studied showed any association with any of the iron status markers (ferritin, serum iron, transferrin, TSAT, sTfR, TIBC or UIBC) or haematological variables (Hb, MCV, MCH, MCHC, RDW, RBC or Hct) or with CRP. However, hepcidin levels varied significantly by rs2235321 genotype (**Figure 2A**) with lower hepcidin in the AA homozygotes, 19% than GG carriers (F ratio 3.70, P=0.014). Note that Bonferroni correction for having analysed 6 SNPs would render the rs2235321 of marginal significance. These trends were stronger in subjects with lower Hb (**Figure 2B**) and lower ferritin levels (**Figures 2C**). The other SNPs had no detectable influence on hepcidin.

Despite the lack of significant association with 5 of the 6 SNPs, we investigated whether allele risk score (ARS) was a significant predictor of hepcidin by using published data on the direction of association to allocate a score of 0, 1 or 2 to allele combinations for each SNP and summing the scores across all SNPs (**Figure 3**).

ANOVA across all ARS revealed no association with plasma hepcidin (F-ratio = 1.01, *p*=0.458).

TF variants

The two *TF* SNPs were also tested in combination and individually. Rs1799852 showed no association with any outcomes. However, rs3811647 was strongly associated with transferrin levels (**Figure 4A**) with or without the inclusion of sex and age as co-variables (F ratio 16.0, P<0.0001). There was an apparent allele dose effect with AA homozygotes having 21% higher transferrin than GG. TIBC (partially computed from transferrin) was similarly affected with a 16% higher value for the AA genotype (F ratio 14.0, P<0.0001). UIBC, which is directly measured rather than computed, showed the same pattern (**Figure 4B**) with 24% higher values in individuals carrying AA (F ratio 12.8, P<0.0001). Serum iron was not significantly associated with the rs3811647 genotype. So, on account of the raised TIBC and transferrin, TSAT was lower in the AA group (by 25%, with a single allele effect of 12.5%) (F ratio 4.3, P<0.0001) (**Figure 4C**). The transferrin, UIBC and TSAT associations were robust when we separated the subjects into above and below the median ferritin value (**Figures 4A, B & C**). None of the other iron markers was affected nor was there any influence on any of the haematological markers, hepcidin or CRP.

4.5. Discussion

Based upon prior GWAS studies, pathway analysis, availability on the Illumina Exome Array, and having a high minor allele frequency in our Gambian population we studied the effect of six candidate SNPs in *TMPRSS6* and two in *TF* on multiple indices of iron and haematological status in 1316 individuals from 1 to 87 years age, from the Keneba Biobank at the MRCG @ LSHTM, in the Gambia. We found weak evidence that one *TMPRSS6* SNP (rs2235321) had lower hepcidin levels in the variant (AA) homozygotes with an indication of an allele dose effect. One *TF* variant (rs3811647) showed increased serum transferrin levels, TIBC and UIBC levels in the AA homozygotes and lower levels of TSAT.

In this population, none of the other variants was significantly associated with any of the iron, haematological, hepcidin or inflammation markers. Applying an allele risk score approach to the 6 *TMPRSS6* variants also yielded no detectable association with any outcomes.

In a separate recall-by-genotype (RbG) study within adults in this same populations, we have assessed fasting hepcidin levels together with iron absorption and changes in plasma hepcidin at two and five hours after an oral dose of 130 mg elemental iron as ferrous sulphate (Jallow et al., in preparation). We recruited individuals homozygous for the variant forms of rs2235321 (n=35) and rs4820268 (n=29) from a panel of 1695 pre-genotyped individuals and compared them to individuals that are wildtype for all of the three variants (n=39). There were no individuals homozygous for the variant allele of rs855791, so we compared heterozygotes (n=28) against the reference group. As in the current study, carriers of the rs2235321 GG genotype had higher fasting hepcidin than AA (GG vs AA, 9.50 vs 6.60ng/ml, $p = 0.035$). But contrary to the current study we also found a difference for rs4820268 (AA vs GG, 9.50 vs

3.27ng/mL, *p*=0.002) and a difference between heterozygotes and wildtype for rs855791 (GG vs AG, 9.50 vs 4.96ng/mL, *p*=0.015). These differences are likely because subjects in the reference group of our recall-by-genotype study were selected as being wildtype at all three SNPs.

Most previous research on the effects of our candidate SNPs were conducted in non-African populations, and there is no prior data on West Africans. The associations we observed differ from results obtained in other populations. *TMPRSS6* rs855791 is a non-synonymous SNP that has been widely reported to influence iron parameters and to be associated with the risk of IDA in Europeans [14,19] and Asians [12,30]. We found no such associations.

The *TMPRSS6* rs2235321 is a synonymous variant which has been reported to associate with benign microcytic anaemia [31]. We did not find any other reported associations. Our data confirm a null effect on hepcidin even in a population with high levels of anaemia and low iron status. A meta-analysis of GWAS on the genetic determinants of hepcidin did not identify any *TMPRSS6* SNP that is significantly associated with hepcidin concentration [13]. It is important to notice that our candidate gene approach demonstrated a relatively small effect of *TMPRSS6* rs2235321 on hepcidin (single allele effect of 9.5%), and this needs to be considered when designing GWAS.

TF rs3811647 is an intron variant on the transferrin gene with extensive prior evidence for functional effects. In discovery and replication GWAS analyses of cohorts from Italy and the USA, Pichler et al. [32] confirmed the association between rs3811647 and transferrin levels. In a subsequent GWAS analysis, Mclaren et al. [19] showed that *TF* rs3811647 is associated with elevated serum TIBC. Also, Blanco-Rojo et al. [21] demonstrated that rs3811647 influenced transferrin gene expression in liver.

Previously, Benyamin et al. showed that three variants in *TF* (rs3811647, rs1799852 and rs2280673) plus the *HFE* C282Y mutation explained ~40% of genetic variation in serum transferrin ($p = 7.8 \times 10^{-25}$) [17]. Our data are suggestive of an allele dose-response relationship with a single allele effect of 9.5% for transferrin levels (higher with the A variant) and a reverse effect of about 12.5% for TSAT.

Other investigators have reported associations between the *TF* rs1799852 and iron status. In a study of female black South Africans, Gichohi and colleagues reported that heterozygotes at *TF* rs1799852 (AG) had lower iron status (low serum ferritin and body iron, and higher sTfR concentrations) than the homozygotes (AA) [23]. This suggested that rs1799852 AA might be protective against low iron status. Similarly, Benyamin and colleagues reported that the *TF* rs1799852 was associated with lower transferrin concentration and the risk of haemochromatosis [33]. Furthermore, Blanco-Rojo and colleagues [18] reported that the *TF* rs1799852 A allele was associated with low serum transferrin concentration, and it compensated for the effect of rs3811647 A allele on the risk of IDA. The authors further suggested that carrying *TF* rs3811467 G allele simultaneously with rs1799852 A allele and *HFE* C282Y and H63D might be protective against low iron status as they increase the susceptibility to iron overload [18]. In the present study, we could not include the *HFE* C282Y and H63D variants, because their MAF in Gambians is extremely low (0.4% and 0% respectively) [34]. Our failure to replicate the prior findings for rs1799852 may be ascribed to the fact that we had only nine individuals homozygous for the A allele. However, in contrast to the South African data [23] , we also found no evidence for differences in any of the iron markers between rs1799852 GG (n=854) and AG (n=117).

This study has strengths and weaknesses. We used highly standardised laboratory assays to measure seven markers of iron status, seven haematological traits, plus

hepcidin and CRP. The sample size was large in the context of candidate gene studies, and we spanned the age range 1-87y (with appropriate adjustment for age and sex in the analyses). The population generally has marginal iron status and high levels of anaemia which might better expose underlying genetic effects. We were limited to the six *TMPRSS6* and two *TF* SNPs available on the exome chip. By definition, these SNPs had been curated onto the chip because of prior evidence of functionality. A limitation is that for some of the SNPs (notably rs1799852 discussed above) we had very few individuals homozygous for the variant allele.

In conducting this study, our initial objective was to explore whether common genetic variants in iron regulatory pathways might have a significant influence on the risk of iron deficiency and iron-deficiency anaemia in African populations. The objective was to provide evidence to inform actionable therapeutic or preventive approaches based on genetic screening. Despite selecting the variants with some strong prior evidence of functionality, our data indicate that such a stratified medicine approach would not be warranted at least in African population. Even where we observed associations (effects of *TMPRSS6* rs2235321 on plasma hepcidin, and *TF* rs3811647 on iron transporting capacity and TSAT), the single allele effect sizes approximated 10%, and there were few people homozygous for the variant allele. Furthermore, there may be inherent compensatory mechanisms because none of the variants had any effect on other markers of iron or haematological status. In summary, the overall population attributable risk conferred by these known genetic factors is negligible. As these variants were mainly studied in European and Asian populations, it is possible that other genetic variants in these genes will be more informative for iron studies in African populations, and this needs to be addressed to develop genetically stratified approaches to prevention or treatment of iron deficiency anaemia.

4.6. NOTES

Acknowledgements

The authors wish to thank Ebrima A. Sise, Alhassan Colley, Ebrima Bah for assisting in the laboratory analysis, Kabiru Sise and the Biobank Field Team for coordinating the participant recruitment for the Biobank Project; Dr Branwen Hennig for setting up the Keneba Biobank and for her expert advice; the Keneba Data Management Team for all of their help.

Funding

The MRC International Nutrition Group is supported by the UK Medical Research Council (grant MC-A760-5QX00) and the UK Department for International Development under the MRC-DFID Concordat agreement. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Conflict of Interests Statement

The authors declare no competing interest

Authors' Contribution

MWJ, AMP and CC designed research; MWJ and CC conducted research; MWJ analyzed the data with input from CC and AMP; MWJ wrote the paper with input from all authors; CC had primary responsibility for final content. All authors read and approved the final manuscript.

Abbreviations used:

GWAS, genome-wide association studies; SNPs, single nucleotide polymorphisms; IRIDA, iron-refractory iron deficiency anaemia; *TMPRSS6*, transmembrane protease serine 6; *TF*, transferrin; HFE, haemochromatosis factor; ARS, allele risk score; FBC, full blood count; RBC, red blood cells; TSAT, transferrin saturation; sTfR, soluble transferrin receptor; UIBC, unsaturated iron binding capacity; TIBC, total iron binding capacity; CRP, C-reactive protein; LD, linkage disequilibrium; EDTA, ethylenedimethyltetraacetic acid; DNA, deoxyribonucleic acid; MCV, mean corpuscular volume; HCT, haematocrit; MCH, mean corpuscular haemoglobin, MCHC; mean corpuscular haemoglobin concentration; ELISA, enzyme-linked immunosorbent assay; MRCG at LSHTM, Medical Research Council Unit the Gambia at London School of Hygiene & Tropical Medicine; NFKBIL1, nuclear factor kB inhibitor–like protein 1.

4.7. References

1. Ganz, T. Systemic Iron Homeostasis. *Physiol. Rev.* **93**, 1721–1741 (2013).
2. Mleczko-Sanecka, K. *et al.* Unbiased RNAi screen for hepcidin regulators links hepcidin suppression to proliferative Ras/RAF and nutrient-dependent mTOR signaling. *Blood* **123**, 1574–1585 (2014).
3. Benyamin, B. *et al.* Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. *Nat. Genet.* **41**, 1173–1175 (2009).
4. Chambers, J. C. *et al.* Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat. Genet.* **41**, 1170–1172 (2009).
5. Li, J. *et al.* GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children. *Hum. Mol. Genet.* **22**, 1457–1464 (2013).
6. Lee, P. Role of Matriptase-2 (TMPRSS6) in Iron Metabolism. *Acta Haematol.* **122**, 87–96 (2009).
7. Heeney, M. M. & Finberg, K. E. Iron-Refractory Iron Deficiency Anemia (IRIDA). *Hematol. Oncol. Clin. North Am.* **28**, 637–652 (2014).
8. Finberg, K. E. *et al.* Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA). *Nat. Genet.* **40**, 569–571 (2008).
9. De Falco, L. *et al.* Iron refractory iron deficiency anemia. *Haematologica* **98**, 845–853 (2013).
10. Ramsay, A. J., Hooper, J. D., Folgueras, A. R., Velasco, G. & Lopez-Otin, C. Matriptase-2 (TMPRSS6): a proteolytic regulator of iron homeostasis. *Haematologica* **94**, 840–849 (2009).

11. Pei, S.-N. *et al.* TMPRSS6 rs855791 Polymorphism Influences the Susceptibility to Iron Deficiency Anemia in Women at Reproductive Age. *Int. J. Med. Sci.* **11**, 614–619 (2014).
12. Bhatia, P., Singh, A., Hegde, A., Jain, R. & Bansal, D. Systematic evaluation of paediatric cohort with iron refractory iron deficiency anaemia (IRIDA) phenotype reveals multiple TMPRSS6 gene variations. *Br. J. Haematol.* **177**, 311–318 (2017).
13. Galesloot, T. E. *et al.* Associations of common variants in HFE and TMPRSS6 with iron parameters are independent of serum hepcidin in a general population: a replication study. *J. Med. Genet.* **50**, 593–598 (2013).
14. Delbini, P. *et al.* Genetic variability of TMPRSS6 and its association with iron deficiency anaemia: Short Report. *Br. J. Haematol.* **151**, 281–284 (2010).
15. Batar, B. *et al.* The role of TMPRSS6 gene variants in iron-related hematological parameters in Turkish patients with iron deficiency anemia. *Gene* **673**, 201–205 (2018).
16. Ji, Y., Flower, R., Hyland, C., Saiepour, N. & Faddy, H. Genetic factors associated with iron storage in Australian blood donors. *Blood Transfus.* (2018) doi:10.2450/2016.0138-16.
17. Benyamin, B. *et al.* Variants in TF and HFE Explain ~40% of Genetic Variation in Serum-Transferrin Levels. *Am. J. Hum. Genet.* **84**, 60–65 (2009).
18. Blanco-Rojo, R. *et al.* Four variants in transferrin and HFE genes as potential markers of iron deficiency anaemia risk: an association study in menstruating women. *Nutr. Metab.* **8**, 69 (2011).
19. McLaren, C. E. *et al.* Associations between Single Nucleotide Polymorphisms in Iron-Related Genes and Iron Status in Multiethnic Populations. *PLoS ONE* **7**, e38339 (2012).

20. Manjari, K. S. *et al.* Transferrin (rs3811647) gene polymorphism in iron deficiency anemia. *Mol. Cytogenet.* **7**, P38 (2014).
21. Blanco-Rojo, R., Bayele, H. K., Srai, S. K. S. & Vaquero, M. P. Intronic SNP rs3811647 of the human transferrin gene modulates its expression in hepatoma cells. *Nutr. Hosp.* **27**, 2142–2145 (2012).
22. McLaren, C. E. *et al.* Genome-Wide Association Study Identifies Genetic Loci Associated with Iron Deficiency. *PLoS ONE* **6**, e17390 (2011).
23. Gichohi-Wainaina, W. N. *et al.* Common Variants and Haplotypes in the TF, TNF-α, and TMPRSS6 Genes Are Associated with Iron Status in a Female Black South African Population. *J. Nutr.* **145**, 945–953 (2015).
24. Gichohi-Wainaina, W. N. *et al.* Associations between Common Variants in Iron-Related Genes with Haematological Traits in Populations of African Ancestry. *PLOS ONE* **11**, e0157996 (2016).
25. Kassebaum, N. J. The Global Burden of Anemia. *Hematol. Oncol. Clin. North Am.* **30**, 247–308 (2016).
26. Hennig, B. J. *et al.* Cohort Profile: The Kiang West Longitudinal Population Study (KWLPS)—a platform for integrated research and health care provision in rural Gambia. *Int. J. Epidemiol.* dyv206 (2017) doi:10.1093/ije/dyv206.
27. Bah, A. *et al.* Serum Hepcidin Concentrations Decline during Pregnancy and May Identify Iron Deficiency: Analysis of a Longitudinal Pregnancy Cohort in The Gambia. *J. Nutr.* **147**, 1131–1137 (2017).
28. Jallow, M. W., Cerami, C., Clark, T. G., Prentice, A. M. & Campino, S. Differences in the frequency of genetic variants associated with iron imbalance among global populations. *PLOS ONE* **15**, e0235141 (2020).
29. R Core Team. *A Language and Environment for Statistical Computing.* (2018).

30. An, P. *et al.* TMPRSS6, but not TF, TFR2 or BMP2 variants are associated with increased risk of iron-deficiency anemia. *Hum. Mol. Genet.* **21**, 2124–2131 (2012).
31. National Center for Biotechnology Information. NM_153609.3(TMPRSS6):c.2217C>T (p.Tyr739=). *ClinVar Genomic variation as it relates to human health* https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000262726.2 (2020).
32. Pichler, I. *et al.* Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels. *Hum. Mol. Genet.* **20**, 1232–1240 (2011).
33. Benyamin, B. *et al.* Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat. Commun.* **5**, 4926 (2014).
34. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

Supplementary Materials

Table S1. Allele configuration of the allele risk score from the 6 TMPRSS6 SNPs

Table S2. Details of the allele configuration for the allele risk score from the 2 TF SNPs

Table S3. A list of all the 94 genotype combinations generated from the 6 TMPRSS6 SNPs

Table S4. The effects of TMPRSS6 allele risk score on hepcidin, controlling for age, sex and CRP

4.9. Figure legends

Fig 1 Linkage disequilibrium analysis between SNPs investigated in this study

Fig 2 Influence of *TMPRSS6* rs2235321 on plasma hepcidin levels

A. All data (GG n=416, GA n=586, AA n=262). ANOVA P for trend = 0.004.
B. Sample divided into high and low Hb (<11.5g/dL). High Hb (GG n=245, GA n=344, AA n=162). ANOVA P for trend = 0.02. Low Hb (GG n=171, GA n=242, AA n=100). ANOVA P for trend = 0.0002.
C. Sample divided above and below median ferritin (<26ng/ml). High ferritin (GG n=247, GA n=313, AA n=155). ANOVA P for trend = 0.0004. Low ferritin (GG n=169, GA n=273, AA n=107). ANOVA P for trend = NS.

Fig 3 Influence of *TMPRSS6* allele risk score on plasma hepcidin

Error bars = standard errors (SE)

Fig 4 Influence of *TF* rs3811647 and rs1799852 on plasma iron binding capacity and TSAT

A. **Transferrin rs3811647** All data (GG n=720, GA n=215, AA n=24). ANOVA P for trend <0.0001. Sample divided above and below median ferritin (<26.9ng/ml). High ferritin (GG n=360, GA n=108, AA n=15). ANOVA P for trend <0.0001. Low ferritin (GG n=360, GA n=107, AA n=11). ANOVA P for trend = <0.0001. **rs1799852** All data (GG n=839, GA n=116, AA n=4). ANOVA P for trend = NS.
B. **UIBC rs3811647** All data (GG n=985, GA n=301, AA n=28). ANOVA P for trend <0.0001. Sample divided above and below median ferritin (<26ng/ml). High ferritin (GG n=474, GA n=147, AA n=16). ANOVA P for trend < 0.0001.

Low ferritin (GG n=511, GA n=154, AA n=12). ANOVA P for trend < 0.0001. **rs1799852** All data (GG n=1139, GA n=116, AA n=9). ANOVA P for trend = NS.

C. **TSAT rs3811647** All data (GG n=720, GA n=215, AA n=24). ANOVA P for trend < 0.0001. Sample divided above and below median ferritin (<26ng/ml). High ferritin (GG n=360, GA n=108, AA n=15). ANOVA P for trend < 0.0001. Low ferritin (GG n=360, GA n=107, AA n=11). ANOVA P for trend < 0.0001. **rs1799852** All data (GG n=839, GA n=116, AA n=4). ANOVA P for trend = NS.

4.10. Figures

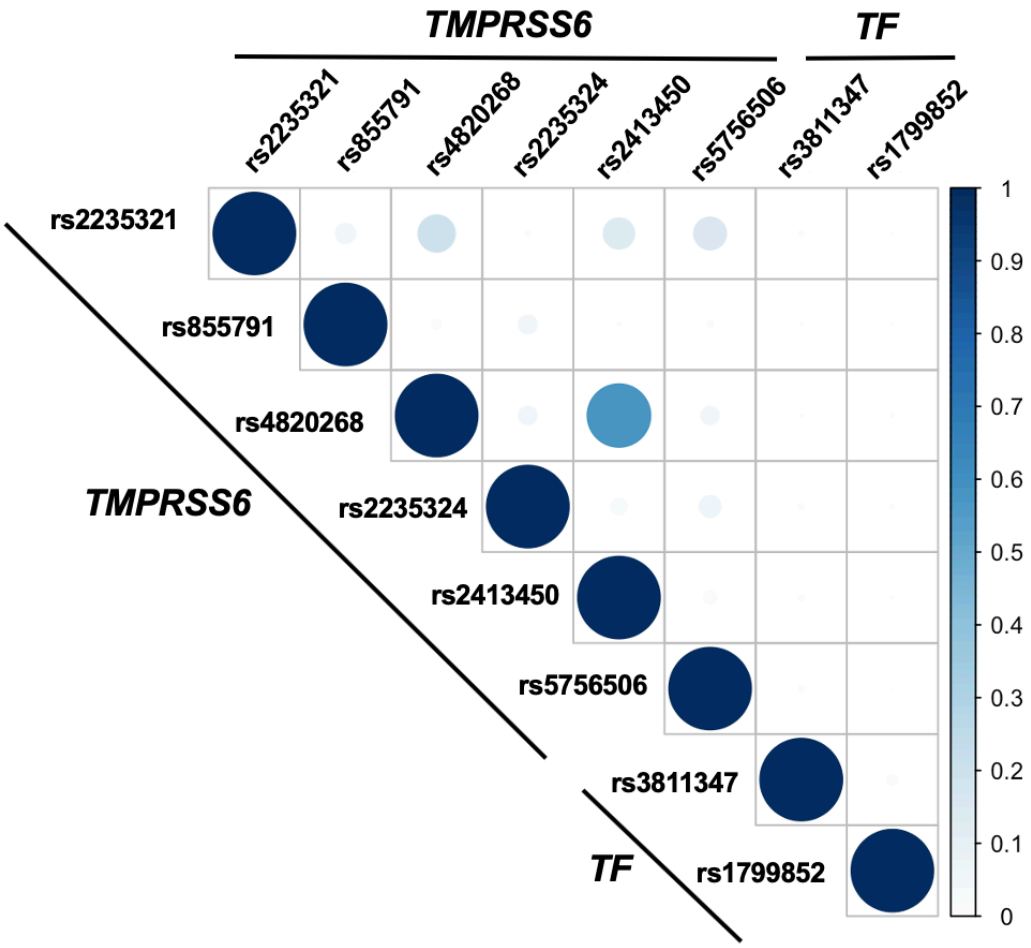


Figure 1

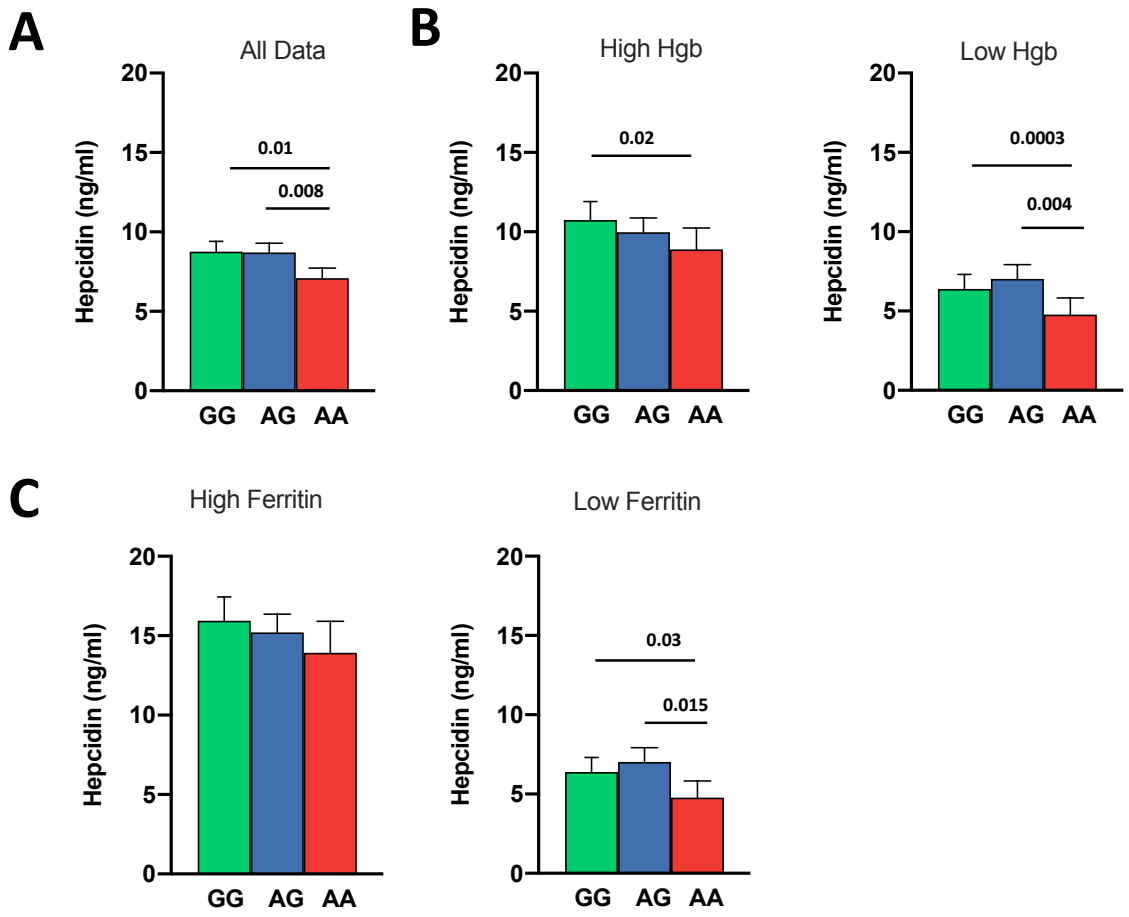

A
All Data
Hepcidin (ng/ml)
20
15
10
5
0
0.01
0.008
GG
AG
AA
B
High Hgb
0.02
Low Hgb
0.0003
0.004
C
High Ferritin
Low Ferritin
0.03
0.015

Figure 2

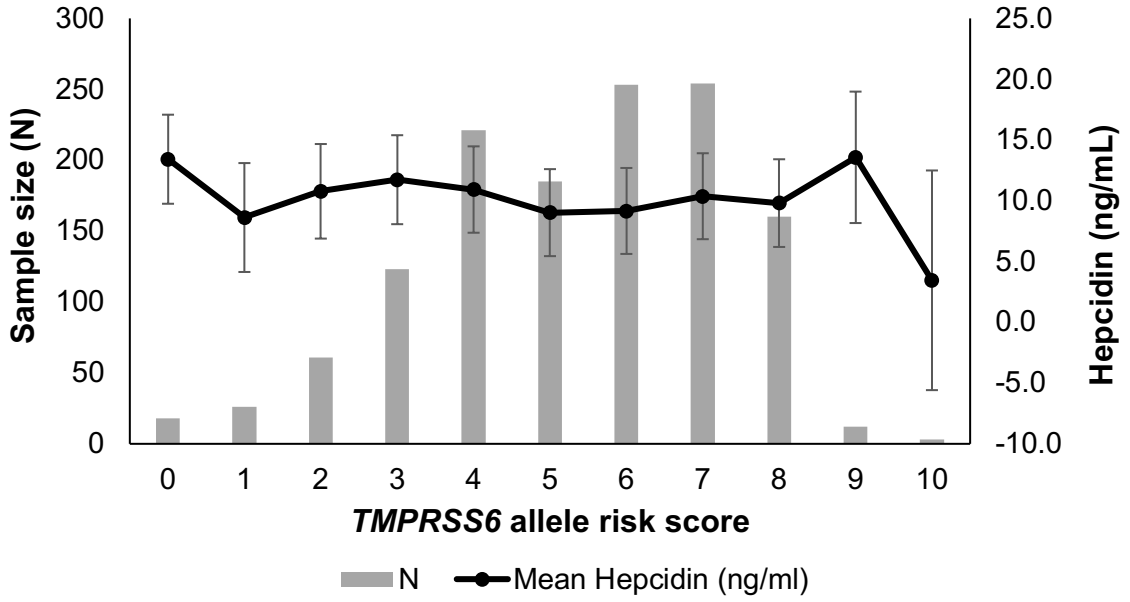

300
250
200
150
100
50
0
Sample size (N)
25.0
20.0
15.0
10.0
5.0
0.0
-5.0
-10.0
Hepcidin (ng/mL)
0 1 2 3 4 5 6 7 8 9 10
TMPRSS6 allele risk score
N
Mean Hepcidin (ng/ml)

Figure 3

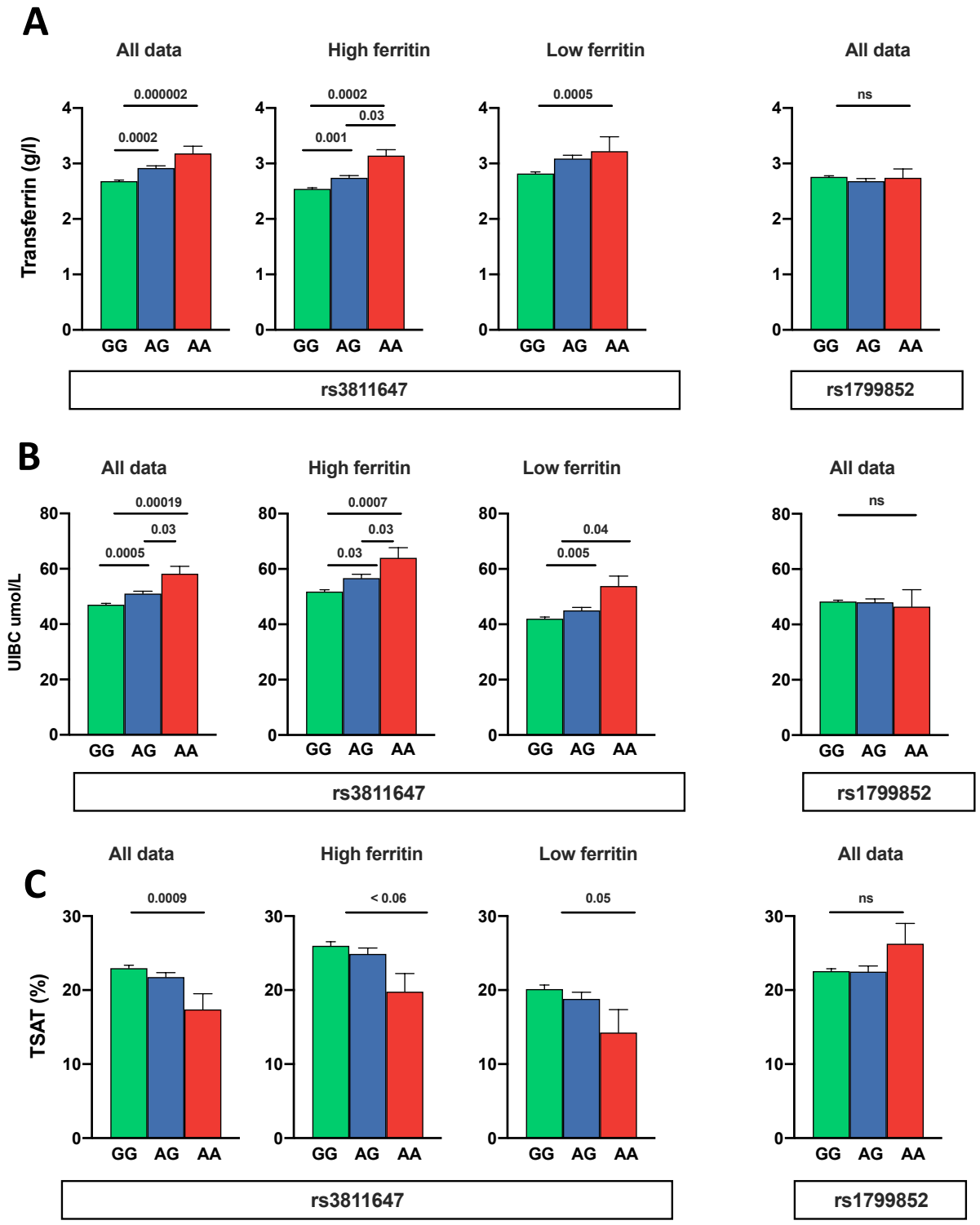

A
All data
High ferritin
Low ferritin
All data
Transferrin (g/l)
0.000002
0.0002
0.0002
0.03
0.001
0.0005
ns
GG AG AA
rs3811647
rs1799852
B
All data
High ferritin
Low ferritin
All data
UIBC umol/L
0.00019
0.03
0.0005
0.0007
0.03
0.03
0.04
0.005
ns
GG AG AA
rs3811647
rs1799852
C
All data
High ferritin
Low ferritin
All data
TSAT (%)
0.0009
< 0.06
0.05
ns
GG AG AA
rs3811647
rs1799852

Figure 4

4.11. Supplemental Data

Association of common *TMPRSS6 and TF* gene variants with hepcidin and iron status in healthy rural Gambians.

Momodou W. Jallow, Susana Campino, Andrew M. Prentice and Carla Cerami

Supplemental information

Supplemental Table S1. Allele configuration of the allele risk score from the 6 *TMPRSS6* SNPs

SNP	rs2235321	rs855791	rs4820268	rs2235324	rs2413450	rs5756506	Total Risk score
major/minor	G/A	G/A	A/G	A/G	G/A	G/C	
Risk allele*	**A**	**A**	**A**	**G**	**G**	**C**	
normal allele	G	G	G	A	A	G	
Risk genotype (ARS)	**A/A (2)**	**A/A (2)**	**A/A (2)**	**G/G (2)**	**G/G (2)**	**C/C (2)**	**12**
normal group	G/G (2)	G/G (0)	G/G (0)	A/A (0)	A/A (0)	G/G (0)	0

ARS, allele risk score

* Allele linked to low iron status from the previously-published studies
Bolded letters are the alleles associated with the risk of low iron status.

Supplemental Table S2. Details of the allele configuration for the allele risk score from the 2 *TF* SNPs

SNP	rs1799852	rs3811647	Genotype combination	ARS	N
major/minor	G/A	G/A			
Risk allele*	**G**	**A**			
WT/WT	**GG**	GG	**GG**/GG	2	825
Het/Het	A**G**	**A**G	A**G**/**A**G	2	10
Homo/Homo	AA	**AA**	AA/AA	2	0
WT/Homo	**GG**	**AA**	GG/AA	4	27
Homo/WT	AA	GG	AA/GG	0	9
Het/WT	A**G**	GG	AG/GG	1	154
WT/Het	**GG**	AG	GG/**A**G	3	290
Het/Homo	A**G**	AA	AG/**AA**	3	1
Homo/Het	AA	AG	AA/AG	1	0

* Allele linked to low iron status from previously-published studies; bolded letters indicates risk alleles

Supplemental Table S3. A list of all the 94 genotype combinations generated from the 6 *TMPRSS6* SNPs

Genotype group	N	*TMPRSS6* SNP ARS
GG/GG/GG/AA/AA/GG	18	0
GG/GG/GG/AA/AG/GG	16	1
GG/GG/GG/GA/AA/GG	10	1
GG/GG/GA/AA/AG/GG	26	2
GG/GG/GG/GA/AG/GG	20	2
AG/GG/GG/AA/AG/GG	4	2
AG/GG/GG/GA/AA/GG	4	2
GG/AG/GG/AA/AA/CG	3	2
GG/GG/GG/GG/AA/GG	3	2
GG/GG/GG/GG/AA/GG	2	2
GG/GG/GG/AA/GG/GG	1	2
GG/GG/GG/GA/AA/CG	1	2
AG/GG/GA/AA/AG/GG	72	3
GG/GG/GA/AA/AG/CG	16	3
GG/GG/GA/AA/GG/GG	16	3
GG/GG/GA/GA/AG/GG	12	3
AA/GG/GG/AA/AG/GG	1	3
AG/GG/GA/NA/AG/NA	1	3
AG/GG/GG/GA/AG/GG	1	3
GG/AG/GA/AA/AG/GG	1	3
GG/AG/GG/GA/AA/CG	1	3
AG/GG/GA/GA/AG/GG	81	4
GG/GG/GA/GA/AG/CG	52	4
AG/GG/GA/AA/GG/GG	29	4
GG/AG/GA/GA/AG/GG	20	4
GG/GG/AA/AA/GG/GG	14	4
GG/GG/GA/GA/GG/GG	9	4
GG/GG/GA/AA/GG/CG	6	4
AG/AG/GG/AA/AG/CG	2	4
AG/GG/GG/GA/GG/GG	2	4
GG/GG/GA/GG/AG/GG	2	4
GG/GG/GG/GG/GG/GG	2	4
AA/GG/GA/AA/AG/GG	1	4
GG/GG/GG/GG/AG/CG	1	4

Supplemental Table S3 cont.

Genotype group	N	*TMPRSS6* SNP ARS
AG/GG/AA/AA/GG/GG	42	5
AG/GG/GA/GA/GG/GG	42	5
GG/GG/GA/GA/GG/CG	25	5
AG/GG/GA/GG/AG/GG	16	5
GG/GG/GA/GG/AG/CG	8	5
GG/GG/AA/GA/GG/GG	7	5
AA/GG/GA/AA/GG/GG	5	5
AG/AG/GA/AA/AG/CG	5	5
GG/AG/GA/GA/AG/CG	5	5
GG/GG/AA/AA/GG/CG	5	5
GG/AG/GA/GA/GG/GG	4	5
AA/GG/GA/GA/AG/GG	3	5
AG/GG/GA/AA/GG/CG	3	5
GG/GG/GA/GG/GG/GG	2	5
AG/AG/GA/GA/AG/GG	1	5
AG/GG/GA/GA/AG/CG	1	5
GG/AG/AA/AA/GG/GG	1	5
GG/AG/GA/AA/AG/CC	1	5
NA/GG/AA/GA/GG/GG	1	5
AA/GG/AA/AA/GG/GG	75	6
AG/GG/AA/GA/GG/GG	56	6
AG/GG/AA/AA/GG/CG	24	6
GG/GG/AA/GA/GG/CG	20	6
GG/AG/AA/GA/GG/GG	19	6
AG/GG/GA/GG/GG/GG	16	6
AA/GG/GA/GA/GG/GG	12	6
GG/GG/GA/GG/GG/CG	7	6
GG/AG/GA/GG/GG/GG	5	6
AA/GG/GA/GG/AG/GG	4	6
AG/AG/GA/GA/AG/CG	3	6
AG/GG/GA/GA/GG/CG	2	6
GG/AG/GA/GA/GG/CG	2	6
GG/GG/AA/AA/GG/CC	2	6
AG/AG/GA/GA/GG/GG	1	6
AG/GG/GA/GG/AG/CG	1	6

Supplemental Table S3 cont.

Genotype group	N	*TMPRSS6* SNP ARS
GG/AG/GA/GA/AG/CC	1	6
GG/GG/AA/GG/GG/GG	1	6
NA/GG/AA/GA/GG/CG	1	6
AA/GG/AA/GA/GG/GG	123	7
AG/GG/AA/GA/GG/CG	62	7
AG/AG/AA/GA/GG/GG	32	7
AG/GG/AA/GG/GG/GG	19	7
GG/GG/AA/GA/GG/CC	8	7
GG/GG/AA/GG/GG/CG	6	7
GG/AG/AA/GA/GG/CG	4	7
GG/AG/AA/GG/GG/GG	2	7
AG/GG/AA/GG/GG/CG	50	8
AA/GG/AA/GG/GG/GG	46	8
AG/AG/AA/GG/GG/GG	30	8
GG/GG/AA/GG/GG/CC	14	8
GG/AG/AA/GG/GG/CG	13	8
AG/AG/AA/GA/GG/CG	6	8
GG/AG/GA/GG/AG/GG	6	8
GG/AA/AA/GG/GG/GG	1	8
GG/AG/AA/GG/GG/CC	6	9
AG/AG/AA/GG/GG/CG	2	9
AG/GG/AA/GG/GG/CC	2	9
AA/GG/AA/GG/GG/CG	1	9
GG/AA/AA/GG/GG/CG	1	9
GG/AA/AA/GG/GG/CC	3	10

Abbreviations: ARS, allele risk score
Legend: NA, genotypes that were not available for that particular SNP in the combination.
TMPRSS6 SNP ARS, number of alleles associated with low iron within a genotyoe combination

Supplemental Table S4. The effects of *TMPRSS6* allele risk score on hepcidin, controlling for age, sex and CRP

TMPRSS6 ARS*	N	Mean Hepcidin (ng/ml)	Std. error	Beta	*P*-value
0	18	13.4	3.66	Reference group	
1	26	8.6	4.48	-4.80	0.285
2	61	10.8	3.89	-2.64	0.498
3	123	11.7	3.66	-1.68	0.645
4	221	10.9	3.55	-2.49	0.483
5	185	9.0	3.58	-4.39	0.220
6	253	9.2	3.54	-4.26	0.229
7	254	10.4	3.54	-3.04	0.390
8	160	9.8	3.61	-3.61	0.317
9	12	13.6	5.41	0.16	0.977
10	3	3.5	9.04	-9.96	0.270

* Number of risk alleles based on **Table S1**

Chapter 5:

A recall-by-genotype study on polymorphisms in the *TMPRSS6* gene and oral iron absorption: a study protocol

Chapter description:

This chapter presents the recall-by-genotype study protocol which is published in F1000 Research.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1513421	**Title**	Mr.
First Name(s)	Momodou W.		
Surname/Family Name	Jallow		
Thesis Title	The impact of single nucleotide polymorphisms in human genes that regulate hepcidin and iron on oral iron absorption and the risk of anaemia in Africans		
Primary Supervisor	Dr Susana Campino		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	**F1000 Research**		
When was the work published?	**May 2019**		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion		This work was published after registration as part of the thesis	
Have you retained the copyright for the work?*	**NO**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be Published?	**N/A**
Please list the paper's authors in the intended authorship order:	**N/A**

Stage of publication	**N/A**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	**I am the first author of this work and I contributed to the design study and write up of the paper**

SECTION E

Student Signature	[redacted]
Date	**13 October 2020**

Supervisor Signature	[redacted]
Date	**13 October 2020**

F1000Research F1000Research 2019, 8:701 Last updated: 02 JUL 2019

STUDY PROTOCOL

A recall-by-genotype study on polymorphisms in the *TMPRSS6* gene and oral iron absorption: a study protocol [version 1; peer review: 2 approved with reservations]

Momodou W. Jallow [1], Susana Campino[2], Andrew M. Prentice[1], Carla Cerami [1]

[1]Nutrition - MRC Unit The Gambia, London School of Hygiene and Tropical Medicine, Fajara, The Gambia
[2]Department of Pathogen Molecular Biology, London School of Hygiene & Tropical Medicine, London, UK

v1 **First published:** 21 May 2019, **8**:701 (https://doi.org/10.12688/f1000research.19080.1)
Latest published: 21 May 2019, **8**:701 (https://doi.org/10.12688/f1000research.19080.1)

Abstract

Background: Oral iron supplementation is commonly used to treat and prevent anaemia. The transmembrane protease serine 6 gene (*TMPRSS6*), which encodes matriptase 2, is a negative regulator of hepcidin, the key controller of iron homeostasis. Genome-wide association studies (GWAS) have identified several single nucleotide polymorphisms (SNPs) in the *TMPRSS6* gene that are associated with an increased risk of iron-deficiency anaemia. We will investigate the *in vivo* effects of three previously reported *TMPRSS6* variants (rs855791, rs4820268 and rs2235321) on oral iron absorption in non-anaemic volunteers in The Gambia.

Methods: A recall-by-genotype study design will be employed. Pre-genotyped participants will be recruited from the West African BioResouce (WABR), which currently contains over 3000 genotyped individuals. Male and female volunteers will be selected based on polymorphisms (rs855791, rs4820268 and rs2235321) in the *TMPRSS6* gene in the Gambian population. The effects of a single variant allele at one SNP and the additive effect of two or three variant alleles from either two or all three SNPs will be investigated. Study participants will be given a single oral dose of 400mg ferrous sulfate, and blood samples will be collected at baseline, two hours and five hours post supplementation. Differences in iron absorption between genotype groups will be assessed by measuring the increase in serum iron concentration at five hours post iron ingestion.

Discussion: This study will increase understanding of the role of genetic variations in *TMPRSS6* on oral iron absorption in subjects of West African origin. This will test for the biological basis for the association of each of the three *TMPRSS6* variants with iron absorption. This may help in guiding future iron intervention strategies, particularly in populations with a high frequency of these SNPs and a high frequency of anaemia.

Study registration: ClinicalTrials.gov NCT03341338 14/11/17.

Keywords

recall-by-genotype, iron supplementation; anaemia; TMPRSS6; hepcidin regulatory genes; genetic variants.

Open Peer Review

Reviewer Status ? ?

	Invited Reviewers	
	1	**2**
version 1 published 21 May 2019	? report	? report

1 **Dale Nyholt**, Queensland University of Technology, Brisbane, Australia
Hamzeh MESRIAN TANHA, QUT, Brisbane, Australia

2 **Alida Melse-Boonstra**, Wageningen University & Research, Wageningen, The Netherlands

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Carla Cerami (ccerami@mrc.gm)

Author roles: Jallow MW: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation; **Campino S**: Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Prentice AM**: Conceptualization, Funding Acquisition, Writing – Review & Editing; **Cerami C**: Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The work was undertaken with support from core funding obtained for the MRC International Nutrition Group from the UK Medical Research Council (MRC) and Department for International Development (DFID) under the MRC/DFID Concordat (PI: Andrew M Prentice).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Jallow MW, Campino S, Prentice AM and Cerami C. **A recall-by-genotype study on polymorphisms in the *TMPRSS6* gene and oral iron absorption: a study protocol [version 1; peer review: 2 approved with reservations]** F1000Research 2019, **8**:701 (https://doi.org/10.12688/f1000research.19080.1)

First published: 21 May 2019, **8**:701 (https://doi.org/10.12688/f1000research.19080.1)

Abbreviations

AGP: alpha-1-acid glycoprotein, CRP: c-reactive protein, EDTA: ethelenediametelenetatraacetic acid, FBC: full blood count, G6PD: glucose-6-phosphate dehydrogenase, GWAS: genome-wide association study, Hb: haemoglobin, IRIDA: iron-refractory iron deficiency anaemia, KWLPS: Kiang West Longitudinal Population Study; LSHTM: London School of Hygiene & Tropical Medicine, MAF: minor allele frequency, MRCG: Medical Research Council The Gambia, SNP: single nucleotide polymorphism, sTfR: soluble transferrin receptor, TMPRSS6: transmembrane protease serine 6, TSAT: transferrin saturation, UIBC: unsaturated iron binding capacity, WABR: West Africa BioResource, WK: West Kiang

Introduction

Despite aggressive implementation of iron supplementation programs, either alone or in combination with food-based supplementation, the prevalence of anaemia remains high in low- and middle-income countries[1,2]. The World Health Organisation (WHO) has set 2050 as a target date by which the current anaemia burden will be reduced by half. In order to achieve this goal, it will be important to identify the major drivers of anaemia.

The transmembrane protease serine 6 gene (*TMPRSS6*), which encodes for matriptase-2, is one of the negative regulators of hepcidin[3], the key iron homeostasis regulator[4]. When serum iron levels are low, matripase-2 suppresses hepcidin expression, allowing more iron from the diet to be absorbed through the intestines into the bloodstream[5,6]. A single nucleotide polymorphism (SNP) in the *TMPRSS6* gene can lead to decreased expression or inactivation of matripase-2[7], which would then lead to inappropriately elevated hepcidin levels, inhibited iron absorption and would thereby result in an increased risk of anaemia[5].

Multiple SNPs in the *TMPRSS6* gene have been linked to iron-refractory iron deficiency anaemia (IRIDA), a hereditary anaemia that is not responsive to oral iron supplementation[8]. In addition, many SNPs in *TMPRSS6* (including rs855791, rs4820268 and rs3345321) have been linked to an increased risk of iron deficiency anaemia (IDA) in genome-wide association studies (GWAS)[9–11]. In Caucasian populations, rs855791 has been reported to be in strong linkage disequilibrium (LD) with rs4820268 (r^2=0.83) and rs2235321 (r^2=0.44)[12]. Similarly, in Asian populations, rs855791 is reported to be in high LD with rs4820268 (r^2=0.65)[12].

The minor allele frequency (MAF) of these SNPs varies between racial and ethnic groups. In African populations, the MAF of rs855791 is lower (10%) than in East Asians (57%), South Asians (54%) and Europeans (39%)[13]. Similarly, the MAF of rs4820268 is lower in Africans (28%) compared to Europeans (42%), whereas, the MAF of rs2235321 in Africans (41%) is similar to that of the European population (42%)[13]. The effects of these SNPs (rs855791, rs4820268 and r2235321) on iron absorption and hepcidin levels in Subsaharan African populations has not been studied.

We hypothesize that the variant alleles at these SNPs may impair iron absorption and may be partially responsible for the disproportionately high anaemia prevalence in sub-Saharan Africa. Here, we propose to investigate effects of these three *TMPRSS6* SNPs on oral iron absorption in Gambian adults.

We anticipate that this study will provide a biological insight into the association of these three *TMPRSS6* variants with anaemia.

Protocol

Study objectives and outcome measures

The primary objective of this study is to assess the impact of single and multiple copies of variant alleles of the *TMPRSS6* SNPs (rs855791, rs4820268 and rs2235321) on oral iron absorption. The primary outcome measure will be the change in serum iron concentration before and five hours after a single 400 mg dose of ferrous sulfate iron given orally (Figure 1).

Secondary endpoints related to the primary objective are:

(1) Increase in transferrin saturation (TSAT) above baseline after a single oral 400 mg dose of ferrous sulfate iron.

(2) Increase in serum unbound iron binding capacity (UIBC) above baseline after a single oral 400 mg dose of ferrous sulfate iron.

(3) Increase in serum hepcidin levels above baseline after a single oral 400 mg dose of ferrous sulfate iron.

(4) Ferritin, haemoglobin, mean corpuscular volume (MCV) and soluble transferrin receptor (sTFR) at baseline, as measures of iron status.

(5) White blood cell count (WBC), granulocyte count, C-reactive protein (CRP) and alpha-1-acid glycoprotein (AGP) at baseline, as measures of the inflammatory state.

(6) Sickle cell haemoglobin and glucose 6-phosphatase deficiency (G6PD) status at baseline to assess potential confounding effects of these two genetic conditions, which are common in this population.

Study design

We will employ a recall-by-genotype study design, in which participant selection will be based on *TMPRSS6* SNPs reported to be associated with the risk of iron-deficiency anaemia: rs855791, rs4820268 and rs2235321[10,14,15]. We will utilize the West African BioResouce (WABR), which contains the Kiang West Longitudinal Population Study (KWLPS) as the basis for selection of pre-genotyped participants[16].

Study site

The proposed study will be conducted within the population of West Kiang (WK) District, in the Lower River Region of The Gambia, and study procedures will be conducted at the Medical Research Council The Gambia (MRCG) at London School of Hygiene & Tropical Medicine (LSHTM), Keneba Field Station[16]. Individuals that are eligible for the study but have moved to the coastal region of The Gambia will be

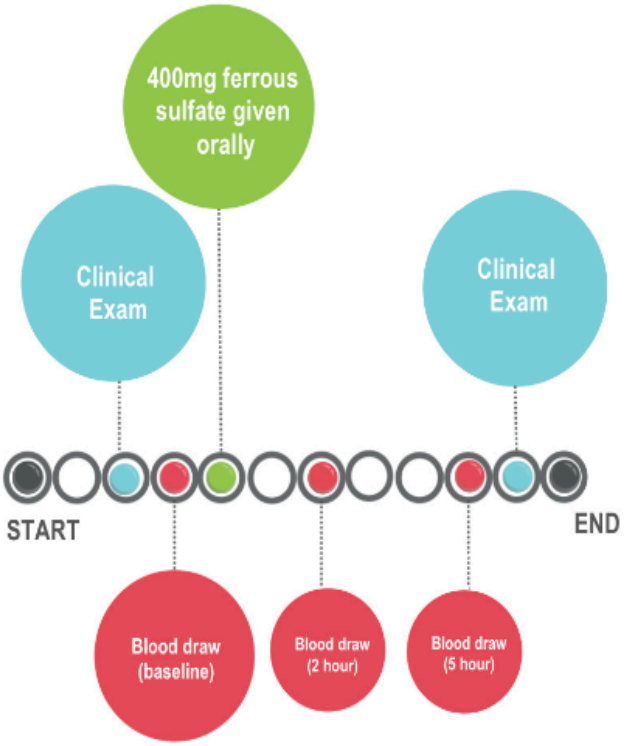


Figure 1. Time line showing oral iron absorption test.

followed-up by a fieldworker and study procedures will be conducted at the MRCG Fajara site. Participants currently residing in WK will be prioritised.

Participants

A total of 300 participants (male and female) will be recruited. Participants will be chosen based on three *TMPRSS6* SNPs (rs855791, rs4820268 and rs2235321), from which we will generate nine genotype combinations, as detailed in Table 1. This will allow the investigation of the effect of each SNP individually and in combination. Composite genotype group 3 is the control group with no variant alleles. Due to the low MAF of rs855791 in our study population, we are unable to include homozygotes for the variant allele. This limited the selection of genotype combinations, and only nine combinations had sufficient participants to include in the study.

For inclusion, participants must be 18 years and above, in good physical health, have available genotype data, be able to fast overnight prior to the study visit and be able to give informed consent. Individuals will be excluded from the study if they have any signs of infection at the time of enrolment, are severely anaemic (Hb <7 g/dl), pregnant or breastfeeding, or have a positive malaria test at screening.

Sample size calculation

The total sample size will be 300. This will include approximately 62 wild type subjects and an average of 31 in each of the eight variant genotype groups. This study size will be able to detect a 12% mean decrease in serum iron at five hours after oral iron supplementation between the wild type and the variant genotype groups with 90% power and a type 1 error of 0 in this study.

Study procedures

Potential participants with the candidate composite genotypes of interest will be selected from the study database by the principal investigator, and contact details (including address and phone number) will be extracted from the WK Demographic Surveillance System[16] by the study data manager. Participants will be contacted either in person or by telephone. Participants who provide informed consent will be invited to the study site where the rest of the study procedures will be conducted, as summarised in Figure 2.

Each participant will be given a single dose of 400mg ferrous sulfate oral iron (2x 200mg ferrous sulfate tablets), equivalent to 130mg elemental iron. To ensure that the iron tablets are taken, a nurse will observe and record the time injestion. Participants will be asked to stay at the study site until the study is completed, which is after collecting the five hour post supplementation blood sample (Figure 1).

All data generated from this study will be anonymised by allocating a unique study ID to each participant. Screening, enrolment and sample collection details will be collected in

Table 1. Genotype combinations based on rs2235321, rs855791 and rs4820268 on the TMPRSS6 gene.

Genotype group	Genotype combination	Rs2235321 wildtype/variant allele	Rs855791 wildtype/variant allele	Rs4820268 wildtype/variant allele	No. of variants alleles
		G/A	G/A	A/G	
1	AA/GG/AA	A/A	G/G	A/A	2
2	AG/GG/GA	A/G	G/G	G/A	2
3	GG/GG/AA	G/G	G/G	A/A	0
4	GG/GG/GA	G/G	G/G	A/G	1
5	GG/GG/GG	G/G	G/G	G/G	2
6	AG/AG/AA	G/A	G/A	A/A	2
7	AG/GG/AA	G/A	G/G	A/A	1
8	GG/AG/AA	G/G	G/A	A/A	1
9	GG/AG/GA	G/G	G/A	A/G	2

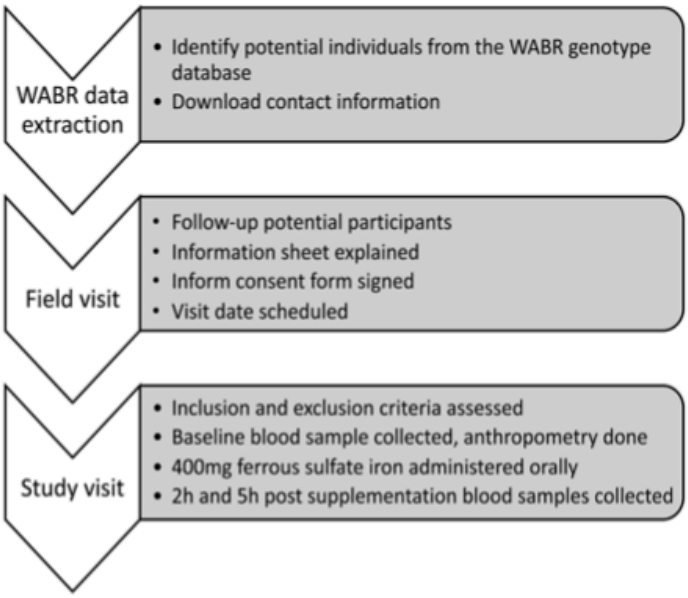


Figure 2. Flow chart showing the study procedures. WABR = West Africa Bioresource.

standard study forms and entered into the study database. Data will be double-entered by two data entry clerks and verified by a data supervisor.

In order to prevent bias in treatment, the composite genotype of individuals will not be disclosed to the study team (data management, field and clinical staff). In addition, participants will be recruited in groups at random, and individuals with different composite genotype groups will be mixed during study visits.

Sample collection

A 3ml whole blood sample will be collected at baseline. 2.5ml will be collected in lithium heparin tubes. 500µl will be collected in EDTA (ethylenediaminetetraacetic acid) micro tubes to be used for full blood count (FBC), malaria rapid testing and sickle screening.

Post supplementation blood samples (3ml blood sample in lithium heparin tube) will be collected at two hours and five hours following iron ingestion. Pre- and post-supplementation

blood samples in lithium heparin tubes will be spun and the plasma aliquoted in barcode-labelled tubes and stored at -20°C for iron biomarker analysis.

Laboratory analyses

FBC will be analysed using a 3-part haematology analyser (Medonic M-series, Boule Medical, Sweden). Iron biomarkers [serum iron, unsaturated iron binding capacity (UIBC), ferritin, soluble transferrin receptor (sTfR), haptoglobin (HP)] and inflammatory markers [C-reactive protein (CRP) and alpha-1-acid glycoprotein (AGP)] will be measured using a Cobas Integra 400 plus biochemistry analyser (Roche Diagnostics). Total iron binding capacity and transferrin saturation of iron (TSAT) will be calculated from serum iron and UIBC. Plasma hepcidin levels will be measured using a commercially available ELISA (DRG Instruments GmbH, Germany). The sickle rapid test will be analysed using the sodium metabisulphide method and positive samples will be genotyped by Hb electrophoresis. G6PD deficiency will be assessed using a qualitative enzyme assay (G6PD Hb+ R&D Diagnostics).

Statistical analysis plan

Primary analysis will be to assess the change in serum iron between the composite genotype groups at the five hours post-supplementation time point. A linear model will be fitted with genotype group as the independent variable and serum iron or TSAT as response variables and genotype group as the main predictor, with the inclusion of age, sex and inflammation status (CRP and AGP levels) as covariates. Using the same approach, we will also examine the effect of genotype on secondary outcome measures. The baseline iron level of the participants may vary. All secondary analysis are exploratory.

In order to remove this potential source of bias, we will adjust for baseline serum iron in the regression analysis. If the missing data rate is more than 5%, we will consider imputation. The follow-up duration is short; thus, we expect little bias from loss to follow-up. We will also consider sensitivity analysis, fitting a multivariate regression model where the main outcomes of interest (including TSAT, iron and hepcidin) will be jointly regressed to the same set of predictors.

Ethical statement

This study has been approved by the MRC Unit The Gambia at the LSHTM Scientific Coordinating Committee, MRC Unit The Gambia at the LSHTM / Gambia Government Joint Ethics Committee (SCC1429), and the LSHTM Ethics Committee (LSHTM Ethics reference number 11679). A trained field worker will visit each potential study participant to issue an information sheet detailing the purpose and nature of the study (see *Extended data*)[17]. Individuals who cannot read will have the information sheet translated into a language they understand by the fieldworker, in presence of an independent witness. Furthermore, participants will be given the opportunity to ask questions to the investigators that they deem important. Participants will be informed that they are free to withdraw from the study anytime, and they can further raise any question about the study with the investigators.

Participants will provide written informed consent, and those who cannot write will provide a thumbprint prior to enrolling into the study. Confidentiality of study participants will be protected by anonymising all study samples and forms by allocating a study number to each participant.

This study was retrospectively registered with ClinicalTrials.gov (NCT03341338) on 14th November 2017.

Dissemination of information

The study results will be published in relevant peer-reviewed journals and key findings will be presented at international scientific meetings. Data sharing will be in agreement with the MRC policy on research data sharing.

Study status

The study is in the data collection phase at the time of publication.

Discussion

GWAS has identified several genetic variants associated with iron status[3,11,15,18–20]. However, detailed understanding of genotype-phenotype relationships is required to identify their effects on iron absorption. The recall-by-genotype (RbG) study design is an efficient tool for detailed investigations of genotype-phenotype relationships because it minimizes confounders and improves statistical power while reducing sample size[21]. In this study, we will use the RbG study design to assess the functional effects of the three common *TMPRSS6* variants on iron absorption. We expect that this study will provide new insights into the association between these *TMPRSS6* gene variants and oral iron absorption in a population where anaemia prevalence is high.

Data availability

Underlying data

No underlying data are associated with this article

Extended data

Figshare: Jallow *et al.* Patient Information sheet and consent form.docx. https://doi.org/10.6084/m9.figshare.8058959.v2[17]

Grant information

The work was undertaken with support from core funding obtained for the MRC International Nutrition Group from the UK Medical Research Council (MRC) and Department for International Development (DFID) under the MRC/DFID Concordat (PI: Andrew M Prentice).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The authors wish to acknowledge Dr. Branwen Hennig for her prior work on the KSWLPS and mentorship, and Dr. Laura Corbin for critically reading the manuscript.

References

1. Kassebaum NJ, GBD 2013 Anemia Collaborators: **The Global Burden of Anemia.** *Hematol Oncol Clin North Am.* 2016; **30**(2): 247–308.
PubMed Abstract | Publisher Full Text
2. Pasricha SR, Drakesmith H: **Iron Deficiency Anemia: Problems in Diagnosis and Prevention at the Population Level.** *Hematol Oncol Clin North Am.* 2016; **30**(2): 309–25.
PubMed Abstract | Publisher Full Text
3. De Falco L, Silvestri L, Kannengiesser C, *et al.*: **Functional and clinical impact of novel *TMPRSS6* variants in iron-refractory iron-deficiency anemia patients and genotype-phenotype studies.** *Hum Mutat.* 2014; **35**(11): 1321–9.
PubMed Abstract | Publisher Full Text
4. Ganz T: **Systemic iron homeostasis.** *Physiol Rev.* 2013; **93**(4) 1721–1741.
PubMed Abstract | Publisher Full Text
5. Finberg KE, Whittlesey RL, Fleming MD, *et al.*: **Down-regulation of Bmp/Smad signaling by Tmprss6 is required for maintenance of systemic iron homeostasis.** *Blood.* 2010; **115**(18): 3817–3826.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Du X, She E, Gelbart T, *et al.*: **The serine protease TMPRSS6 is required to sense iron deficiency.** *Science.* 2008; **320**(5879): 1088–92.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Silvestri L, Pagani A, Nai A, *et al.*: **The serine protease matriptase-2 (TMPRSS6) inhibits hepcidin activation by cleaving membrane hemojuvelin.** *Cell Metab.* 2008; **8**(6): 502–511.
PubMed Abstract | Publisher Full Text | Free Full Text
8. De Falco L, Sanchez M, Silvestri L, *et al.*: **Iron refractory iron deficiency anemia.** *Haematologica.* 2013; **98**(6): 845–53.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Sal E, Keskin EY, Yenicesu I, *et al.*: **Iron-refractory iron deficiency anemia (IRIDA) cases with 2 novel *TMPRSS6* mutations.** *Pediatr Hematol Oncol.* 2016; **33**(3): 226–32.
PubMed Abstract | Publisher Full Text
10. Pelusi S, Girelli D, Rametta R, *et al.*: **The A736V TMPRSS6 polymorphism influences hepcidin and iron metabolism in chronic hemodialysis patients: TMPRSS6 and hepcidin in hemodialysis.** *BMC Nephrol.* 2013; **14**: 48.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Delbini P, Vaja V, Graziadei G, *et al.*: **Genetic variability of *TMPRSS6* and its association with iron deficiency anaemia.** *Br J Haematol.* 2010; **151**(3): 281–284.
PubMed Abstract | Publisher Full Text
12. Chambers JC, Zhang W, Li Y, *et al.*: **Genome-wide association study identifies variants in *TMPRSS6* associated with hemoglobin levels.** *Nat Genet.* 2009; **41**(11): 1170–2.
PubMed Abstract | Publisher Full Text | Free Full Text
13. 1000 Genomes Project Consortium, Auton A, Brooks LD, *et al.*: **A global reference for human genetic variation.** *Nature.* 2015; **526**(7571): 68–74.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Gichohi-Wainaina WN, Tanaka T, Towers GW, *et al.*: **Associations between Common Variants in Iron-Related Genes with Haematological Traits in Populations of African Ancestry.** *PLoS One.* 2016; **11**(6): e0157996.
PubMed Abstract | Publisher Full Text | Free Full Text
15. An P, Wu Q, Wang H, *et al.*: ***TMPRSS6*, but not *TF*, *TFR2* or *BMP2* variants are associated with increased risk of iron-deficiency anemia.** *Hum Mol Genet.* 2012; **21**(9): 2124–2131.
PubMed Abstract | Publisher Full Text
16. Hennig BJ, Unger SA, Dondeh BL, *et al.*: **Cohort Profile: The Kiang West Longitudinal Population Study (KWLPS)-a platform for integrated research and health care provision in rural Gambia.** *Int J Epidemiol.* 2017; **46**(2): e13.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Jallow MW, Campino S, Cerami C, *et al.*: **Jallow et al Patient Information sheet and consent form.docx.** 2019.
http://www.doi.org/10.6084/m9.figshare.8058959.v2
18. Benyamin B, Ferreira MA, Willemsen G, *et al.*: **Common variants in *TMPRSS6* are associated with iron status and erythrocyte volume.** *Nat Genet.* 2009; **41**(11): 1173–5.
PubMed Abstract | Publisher Full Text | Free Full Text
19. McLaren CE, Garner CP, Constantine CC, *et al.*: **Genome-wide association study identifies genetic loci associated with iron deficiency.** *PLoS One.* 2011; **6**(3): e17390.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Nai A, Pagani A, Silvestri L, *et al.*: **TMPRSS6 rs855791 modulates hepcidin transcription *in vitro* and serum hepcidin levels in normal individuals.** *Blood.* 2011; **118**(16): 4459–4462.
PubMed Abstract | Publisher Full Text
21. Corbin LJ, Tan VY, Hughes DA, *et al.*: **Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference.** *Nat Commun.* 2018; **9**(1): 711.
PubMed Abstract | Publisher Full Text | Free Full Text

Chapter 6:

Common variants in the transmembrane protease serine 6 (*TMPRSS6*) gene alters hepcidin but not plasma iron in response to oral iron in healthy Gambian adults: a recall-by-genotype study

Chapter description:

This chapter presents the results of the recall-by-genotype study on the effects of common *TMPRSS6* SNPs on oral iron absorption.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1513421	**Title**	Mr.
First Name(s)	Momodou W.		
Surname/Family Name	Jallow		
Thesis Title	The impact of single nucleotide polymorphisms in human genes that regulate hepcidin and iron on oral iron absorption and the risk of anaemia in Africans		
Primary Supervisor	Dr Susana Campino		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	N/A		
When was the work published?	N/A		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be Published?	**Current Developments in Nutrition**
Please list the paper's authors in the intended authorship order:	**Momodou W. Jallow, Susana Campino, Alasana Saidykhan, Andrew M Prentice and Carla Cerami**

Stage of publication	**Under revision**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	**This is the recall-by-genotype study. I contributed to conceptualisation and design of this study. I led the date collection and analysis and I drafted the manuscript, managed co-author comments, and the submission process.**

SECTION E

Student Signature	[redacted]
Date	**15 January 2021**

Supervisor Signature	[redacted]
Date	**15 January 2021**

Common variants in the transmembrane protease serine 6 (*TMPRSS6*) gene alter hepcidin but not plasma iron in response to oral iron in healthy Gambian adults: a recall-by-genotype study

Short title: Effects of *TMPRSS6* variants on plasma iron

Momodou W. Jallow[1,2], Susana Campino[2], Alasana Saidykhan[1], Andrew M. Prentice[1] and Carla Cerami[1*]

Author Affiliations

[1] MRC Unit The Gambia at London School of Hygiene & Tropical Medicine, Atlantic Boulevard, Fajara, P.O. Box 273, Banjul, The Gambia
[2] Department of Infection Biology , Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

Corresponding author:
Dr Carla Cerami, MRC Unit The Gambia at London School of Hygiene & Tropical Medicine, Atlantic Boulevard, Fajara, P.O. Box 273, Banjul, The Gambia
Email: ccerami@mrc.gm Tel: +220 7875756.

List of abbreviations
AGP, alpha-1-glycoprotein; BMI, body mass index; CRP, C-reactive protein; DNA, deoxyribonucleic acid; EDTA, ethylenedimethyltetraacetic acid; ELISA, enzyme-linked immunosorbent assay; FBC, full blood count; GWAS, genome-wide association studies; G6PD, glucose-6-phosphase dehydrogenase; Hb, Hemoglobin; HCT, hematocrit; *HFE*, hemochromatosis factor; IRIDA, iron-refractory iron deficiency anemia; LD, linkage disequilibrium; LMIC, low- and middle-income countries; MAF; minor allele frequency; MCH, mean corpuscular hemoglobin, MCHC; mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RBC, red blood cells; SNPs, single nucleotide polymorphisms; sTfR, soluble transferrin receptor; *TF*, transferrin; TIBC, total iron binding capacity; *TMPRSS6*, transmembrane protease serine 6; TSAT, transferrin saturation; UIBC, unsaturated iron binding capacity.

Funding
Momodou W. Jallow, Andrew M. Prentice, Carla Cerami and Alasana Saidykhan were supported by the UK Medical Research Council (grant MC-A760-5QX00) and the UK Department for International Development under the MRC-DFID Concordat agreement. Susana Campinno is supported by funding from the MRC UK (MR/R020973/1) and the BBSRC UK (BB/R013063/1).

Conflict of interests
The authors declare no conflict of interest

6.1. Abstract

Background: The role of genetic determinants in mediating iron status in Africans is not fully understood. Genome-wide association studies in non-African populations have revealed genetic variants in the *TMPRSS6* gene that are associated with the risk of anemia.

Objectives: To investigated the effects of risk alleles for low iron status from *TMPRSS6* rs2235321, rs855791 and rs4820268, on responses to oral iron in healthy Gambian adults.

Methods: Using a recall-by-genotype design, participants were selected from a pre-genotype cohort of 3000 individuals in the Keneba Biobank (MRCG at LSHTM). Participants were invited to participate in the study based on nine genotype combinations obtained from three *TMPRSS6* SNPs (rs2235321, rs855791 and rs4820268). The participants fasted overnight and then ingested a single oral dose of ferrous sulfate (130 mg elemental iron). Blood samples were collected prior to iron ingestion and at 2 and 5 hours after the oral iron dose. The effects of genotype on hepcidin and plasma iron parameters were assessed.

Results: A total of 251 individuals were enrolled. Homozygous carriers of the major *TMPRSS6* alleles at each of the SNPs had higher plasma hepcidin at baseline (rs2235321: GG vs AA 9.50 vs 6.60ng/ml, $p = 0.035$; rs855791: GG vs AG = 9.50 vs 4.96ng/mL, $p=0.015$; rs4820268: AA vs GG = 9.50 vs 3.27ng/mL, $p=0.002$) and at subsequent timepoints. There were no differences in delta plasma iron (a proxy for iron absorption) between genotypes. In most subjects, hepcidin levels increased following iron ingestion (overall group mean = 4.98±0.98ng/ml at 5h, $p<0.001$), but double heterozygotes at rs2235321 and rs855791 showed no increase (0.36±0.40ng/ml at 5h, $p=0.667$).

Conclusions: This study revealed that common *TMPRSS6* variants influence hepcidin concentrations, but not iron status indicators either at baseline or following a large oral dose of iron. These results suggest that genetic variations in the *TMPRSS6* gene are unlikely to be important contributors to variations in iron status in Africans. This study was registered at ClinicalTrials.gov # NCT03341338.

Keywords: Anemia, *TMPRSS6*, recall-by-genotype, hepcidin, genetic polymorphism, iron absorption.

6.2. Introduction

Iron supplementation remains the dominant strategy for the prevention and treatment of anemia (1,2). However, despite decades of implementing this measure together with food-based approaches, the prevalence of anemia remains high particularly among children and women of reproductive age living in low- and middle-income countries (LMICs) (3).

Matriptase-2 protein encoded by the transmembrane protease serine 6 (*TMPRSS6*) gene is a negetive regulator of hepcidin, the regulator of iron metabolism (4). Hepcidin blocks the release of intracellular iron by downregulating ferroportin, the only known mammalian cellular iron transporter (5). These effects are especially pronounced in enterocytes and in macrophages and thus elevated hepcidin is associated with impaired duodenal iron absorption and impaired recycling of aged red blood cells (5). Single nucleotide polymorphisms (SNPs) in *TMPRSS6* can lead to decreased function or inactivation of matriptase-2, thus impairing its suppression of hepcidin (*HAMP*) gene expression (6). This leads to inappropriately elevated hepcidin which, by blocking iron absoption and recycling, promotes the risk of iron deficiency and anemia (7). Genome-wide association studies (GWAS) have revealed numerous common SNPs in *TMPRSS6* that are linked to an increased risk of iron deficiency anemia (IDA) (8).

Three single nucleotide polymorphims (rs855791, rs2235321 and rs4820268) have been reported to be associated with low iron status, but mainly in non-African populations (9–12). However, the effects of these SNPs on the response to iron supplementation have not previously been described in Africans. The minor allele frequency (MAF) of these SNPs in Africans in the 1000 Genomes project are 10%, 41% and 28% for rs855791, rs2235321 and rs4820268, respectively (13).

TMPRSS6 rs855791 is a non-synonymous SNP that alters matriptase 2 protein (14), but rs2235321 and rs4820268 are synonymous variants whose direct effect are not clear (15). Although synonymous changes were previously not though to directly affect phenotype, recent findings show that they can affect protein folding and splicing (16). In Caucasians, rs855791 is reported to be in high linkage disequilibrium (LD) with rs4820268 and rs2235321 (9,11), and is in high LD with rs4820268 in Asians (17). However, low LD has been observed between these SNPs in the Africans included in the 1000 Genomes project and in Gambians (18).

In this study, we sought to assess the effects of the three common *TMPRSS6* SNPs, either individually or combined, on the response to a high dose of oral iron in healthy Gambian adults.

6.3. Materials and Methods

Study Design

The full details of the study design were published in the study protocol (19), and the study was registered at ClinicalTrials.gov (NCT03341338). Using a recall-by-genotype approach, participants were enrolled based on their *TMPRSS6* rs2235321 (MAF=43%), rs855791 (MAF=7%) and rs4820248 (MAF=27%) genotypes. We selected participants from the Keneba BioBank at MRCG@LSHTM, which contained 3116 pre-genotyped participants. Out of these 3116 individuals, n=1695 met the criteria for for inclusion in the present study, **Figure 1**.

Genotyping

BioBank participants were previously genotyped using the Infinium 240k Human Exome Beadchip v1.0 and v1.1 (Illumina, CA, USA) . Genotype calling was done using

data-driven clustering (Genome Studio, Illumina, CA, USA). The *TMPRSS6* rs2235321, rs855791 and rs4820268 SNPs were selected based on their previously published associations with measures of iron status.

Genotype combinations

We constructed genotype combinations for each participant from the three candidate *TMPRSS6* SNPs. This generated a total of 17 genotype combinations, **Table 1**. Only nine of these combinations had a sufficient number of individuals to perform grouped analysis. We focused on genotype groups that contained more than 95 individuals.

Participant selection

Participants were selected based on the combinations obtained from combining the 3 possible genotypes for each of the three *TMPRSS6* SNPs studied (**Figure 1**). Individuals were invited to participate if they were between 18 and 50 years of age. Women were excluded if they were breastfeeding or pregnant. Also, individuals who reported to be unwell and those with severe anemia (Hb<7g/dl) were excluded. Individuals that tested positive for malaria were to be excluded but there were none.

Study procedures

Contact details for the potential participants were retrieved from the Kiang West Demographic Database (20). A field worker initially contacted each participant in person or by telephone. Individuals who agreed to participate provided written informed consent (see below) and were invited to the study sites at MRCG Keneba or Fajara, for the investigative procedures.

A baseline blood sample (3mL, 2.5mL in lithium heparin and 0.5mL in EDTA tubes) was taken following an overnight fast. Thereafter, a single dose of 400mg (2x 200mg) ferrous sulphate oral iron, containing 130mg elemental iron, was given by a study nurse. The choice of 400 mg ferrous sulphate was based on the studies by Hwang et al., 2011 [23] which assessed the effects of a high dose of oral iron (650 mg ferrous sulphate) on hepcidin and Nai et al. 2011 [18] which examined the effects of *TMPRSS6* rs855791 on the response to an oral iron dose. However, we reduced the 650 mg ferrous dose to 400 mg, to minimise the iron overload. The high dose was used to elicit a transient plasma iron overload.

Participants were observed to ensure that the supplements were taken and the time of ingestion was recorded. Participants were asked to stay at the study site, and blood samples (2.5mL lithium heparin tubes, at each timepoint) were taken at 2- and 5-hours following iron supplementation. In addition, the weight and height of each participant to enable calculation of body mass index (BMI), and body temperature to assess possible fever, were measured. Plasma iron at the five hour post iron ingestion sample was used as the primary outcome variable and as a proxy for iron absorption. Differences in hepcidin concentration between genotypes was a secondary outcome.

Laboratory procedures

Full blood count (FBC) (Medonic M-Series, Boule Medical, Sweden), malaria rapid test (SD BioLine Malaria Antigen Pf, Standard Diagnostics Inc., Republic of Korea), sickling test (sodium metabisulphide method and Hb electrophoresis for confirmation of Hb genotype for positive samples) and G6PD screening (G6PD Hb+ R&D Diagnostics) were performed on the EDTA sample. The lithium heparin samples were spun, and the plasma stored at -20^{o}C for iron biomarker analysis. Plasma iron, ferritin,

unsaturated iron binding capacity (UIBC), soluble transferrin receptor (sTfR), C-reactive protein (CRP) and alpha-1-glycoprotein (AGP) were measured using an automated biochemistry analyser (Cobas Integra 400 Plus, Roche Diagnostics). Total iron binding capacity (TIBC) and transferrin saturation (TSAT) were calculated from UIBC and iron (TIBC = UIBC+iron and TSAT = [iron/TIBC]*100). For all the biochemistry analysis, the analyser was calibrated using commercial calibrators and controls were analysed for each parameter.

Hepcidin was quantified using a commercial enzyme-linked immunosorbent assay (ELISA) (DRG Instruments GmbH, Germany) according to manufacturer's protocol. To ensure quality of the results, two manufacturer-supplied controls (high and low controls) were analysed alongside the samples in each ELISA plate.

Statistical analysis

Student's *t*-test and the Wilcoxon test were used to determine the differences in iron biomarkers between sexes for non-skewed and skewed data respectively. Chi-square was used to test the differences between categorical data. Linear regression models were used to assess the effects of genetic variants (individual SNPs or genotype combination) on iron markers at each timepoint. The control group (group with individuals homozygous for the major allele at each SNP) was set as the reference in the model. Sex and inflammation status (CRP) were included as covariates to account for their known influence on iron status (21,22). Also, G6PD and sickle cell anemia were accounted for. Furthermore, account for the influence of baseline iron on the response to the oral iron dose, baseline ferritin was included as a covariate in the analysis. Skewed data was log-transformed. Bonferroni correction was applied to

adjust for multiple testing. Statistical analyses were conducted using the R statistical software (23).

Ethics statement

This study was approved by the Medical Research Council Unit The Gambia (MRCG) Scientific Coordinating Committee and the MRCG at London School of Hygiene & Tropical Medicine (LSHTM)/Gambia Government Joint Ethics committee (SCC1429), and the LSHTM Ethics Committee (11679). A fieldworker administered a copy of the study information sheet to each participant. Individuals who could not read had it translated to a language they understood in the presence of an independent witness. Each participant provided a written informed consent prior to enrolling into the study, and those who cannot write provided a thumbprint. To ensure confidentiality of the participants, all the samples and forms were anonymized by allocating study numbers.

6.4. Results

A total of 251 individuals were enrolled in the study, and the number of individuals enrolled in each genotype group is shown in **Table 2**. Due mostly to outward migration of males we had more females (76%). The World Health Organisation (WHO) hemoglobin (Hb) cut-offs (<12.0 g/dL and <13.0 g/dL, non-pregnant women and men, respectively) was applied to determine anemia in the study population. Eighty (31.9%) of the participants were anemic, and 54 (64%) of these were females. From the 80 anemic individuals, 30 (37.5%) were iron deficient (had IDA) (ferritin <15 µg/L and CRP <5.0 g/L), with an overall 11.9% of the study population iron deficient. Twenty-six (86.7%) of the iron deficient individuals were female. The study was conducted in the Kiang West District of the Gambia, which is a rural area about 200 kilometers from

the Capital City (Banjul). Due to the low economic activities in this area, it is common for men to leave for the urban areas in search of employment. The baseline characteristics of the study participants are presented in **Table 3.**

The effects of genotype on plasma iron concentrations pre- and post-iron ingestion

Plasma iron increased significantly in all the genotype groups after the iron dose (**Figure 2**). No significant differences were observed between the genotypes of individual SNPs (**Figure 3A**) and between double heterozygotes and the reference group, in plasma iron both before and after the iron dose (**Figure 3B**).

The effects of genotype on hepcidin concentrations pre- and post-iron ingestion

There were significant differences between genotypes of individual SNPs on hepcidin (**Figure 4A**). For each of the SNPs, carriers of the homozygous major alleles (rs2235321 GG, rs855791 GG and rs4820268 AA, **Figure 4A**) had higher hepcidin concentrations than individuals with the minor alleles, both before and after the iron dose. In addition, when comparing double heterozygotes and the reference group, the latter had the highest hepcidin concentrations at all the timepoints, and this significantly differed from the genotype group AG/AG/AA (**Figure 4B**). The genotype group AG/AG/AA had the lowest hepcidin concentration both before and after the iron dose (**Figure 4B**).

There was an increase in hepcidin concentration in all the genotype groups, following the iron dose (**Figure 5A**), except in one group (AG/AG/AA: double heterozygotes at rs2235321 and rs855791) (**Figure 5B**). The individuals in the genotype group AG/AG/AA had the lowest hepcidin concentrations at baseline, and this remained

unchanged five hours after iron ingestion (**Figure 5B**). Also, carriers of the genotype rs2235321 AA increased their hepcidin concentrations after the iron dose, but the difference between the baseline and five hours was not statistically significant (*P*=0.060) (**Figure 5A**).

The efffects of genotype on TSAT, TIBC, UIBC, ferritin, sTfR, transferrin and hematology traits

There were significant differences between genotypes of each of the SNPs in TIBC and UIBC at baseline, but these differences were not detected at 2- and 5-hours after the iron dose (**Supplemental Table 1)**. There were no significant differences between the genotypes for any of the SNPs in TSAT, transferrin, ferritin or sTfR either before or after iron ingestion (**Supplemental Table 1**). Furthermore, there were no differences between genotypes for any of the SNPs on any of the hematological traits (**Supplemental Table 2**). There were no significant differences between the double heterozygotes and the reference group in TSAT, TIBC and UIBC.

6.5. Discussion

We used a candidate genotype approach to recall individuals with variant alleles of three *TMPRSS6* SNPs previously associated with iron imbalances. We hypothesised that carriers of risk alleles previously reported to be associated with low iron status would have inappropriately elevated hepcidin levels and thus impair response to oral iron.

Our study participants were healthy individuals, and we used the increase in plasma iron at five hours after ingestion of the iron dose as a proxy to measure response to oral iron. We found that all subjects increased their plasma iron concentrations and

TSAT levels at five hours, but there were no differences between genotypes individually or in combination. Therefore, we could not establish that carriage of low-iron risk alleles from any one of these SNPs impairs the response to oral iron.

The *TMPRSS6* rs855791 A allele has been widely associated with IRIDA in Caucasians and Asians (25–28). Similarly, the rs2235321 A and rs4820268 G alleles have been linked to the risk of iron deficiency, including in African populations (10,29,30). In a meta-analysis, Gichohi-Wainaina and colleagues (8) reported the rs855791 A allele to be associated with decreased Hb and ferritin concentrations across all the populations they studied. Therefore, we expected these SNPs to have an effect on plasma iron biomarkers either at baseline or on the response to the iron dose. However, as with the plasma iron, we did not find any effects of these SNPs on ferritin, TSAT or any other iron biomarker either before or after iron ingestion. Also, there were no differences in hematological traits in our study.

In most subjects plasma hepcidin showed the anticipated acute rise in response to the administered iron dose. For rs855791, we observed an unexpected result where, at baseline, GG carriers had higher hepcidin concentrations compared to AG. The same trend was observed at five hours post iron ingestion. This contradicts what has been reported about this SNP in other populations. The rs855791 AA (homozygous for the minor allele) has been associated with elevated hepcidin concentrations accompanied by decreased TSAT and serum iron in Europeans (9). In the study conducted by Nai and colleagues (9), hepcidin decreased in a dose-dependent manner with rs855791 AA having higher hepcidin concentrations than AG and GG carriers. Our results contradict this finding, as we observed carriers of rs855791 GG to have higher hepcidin concentration than AG. However, due to its low MAF, we were unable to include rs855791 AA in the study. Most of the previous reports on rs855791 were

obtained from studies of non-African populations. Hence, our results suggest that rs855791 may have a different effect on hepcidin levels against the different genetic background of West Africans.

There have been a number of recent studies on the effects of *TMPRSS6* variants on iron status in different populations. A study of Pakistani women of reproductive age found that rs855791 T allele (A on the reverse strand) is associated with the risk of IDA (31). However, in the present study, we did not have individuals with the T allele. In a study of South African chronic kidney disease patients, Nalado and colleagues (32) found that the rs855791 C alleles do not predispose to IDA. This finding is similar to our results on this SNP, as we did not find any effects of the major allele (G, opposite strang of C) on iron status indices. Also, in an iron absorption study on Taiwanese women using stable iron isotopes, Buerkli et al., (33) reported that the *TMPRSS6* rs855791 alters iron absorption, and that the carriers of the C alleles absorbed iron better than T allele carriers. Our study was short (5 hours) and we did not use stable isotopes, and hence may not be sensitive for assessing iron absorption.

From the analysis of genotype combinations, we did not find any group that differed significantly from the reference (GG/GG/AA: rs2235321 GG, rs855791 GG and rs4820268 AA) in plasma iron concentrations or TSAT levels. However, we observed that carriers of the genotype combination AG/AG/AA (simultaneous carriage of rs2235321 AG rs855791 AG and rs4820268 AA) maintained a low mean hepcidin concentration up to five hours, despite exposure to a high iron dose. The plasma iron levels of this group rose significanly by five hours post iron ingestion, while the hepcidin levels remained constant. This is in contradiction to the clear rises in hepcidin levels shown for most subjects in this study, and widely reported in the literature (34,35). An acute rise hepcidin levels in response to oral iron is the normal feedback

mechanism to halt further iron absorption when optimal levels are reached (36). In IDA, low levels of hepcidin promote iron absorption, but this is still accompanied by acute elevation of hepcidin concentration (5). Further studies of why this group lacks the acute hepcidin response, so clearly evident in the other subjects, warrants further investigation and might provide insights into the response of hepcidin.

A strength of this study is that, using a recall-by-genotype strategy, we concentrated on informative individuals by focusing on known carriers of our genotypes of interest. This is an efficient method to identify genotype-phenotype relationship with improved statistical power and to eliminate the effects of confounders (37). A major limitation of our study was the low MAF of the variant most widely decribed in the literature (rs855791), and this prevented us from studying homozygotes for the minor allele. Other limitations include the use of high iron dose, which is double the routine dosage, and may be less sensitive for assessing iron absorption. Also, subjects were relatively iron replete with high hepcidin levels which might inhibit iron absorption (36). We therefore cannot exclude the possibility that the gene variants studied may have differential effects under conditions of iron deficiency. Likewise it is possible that responses to a lower (more physiological) dose of iron might differ to the results reported here. Using the post-prandial change in plasma iron as the primary outcome has strengeths and weaknesses. It permitted large numbers of measurements to be made and is a relevant measure of the acute response to iron administration but is only a proxy measure of intestinal iron absorbption (38). As in any study of this type the ability to detect genotype effects will be blunted by natural variance caused by differences in diet, nutritional status of other micronutrients potentially affecting iron status, and potential effects of inflammation. Also, epistatic influences of other genes cannot be excluded.

From this study, we conclude that common *TMPRSS6* variants influence hepcidin, but not post-prandial iron status (a proxy for oral iron absorption). It therefore seems unlikely that genetic variations in the *TMPRSS6* gene are important contributors to differences in iron status in this, and likely other, African pupulation.

6.6. NOTES

Acknowledgement

We thank Alhassan Colley, MRCG at LSHTM Keneba Laboratory, for conducting the biochemistry analysis; Bakary Sonko and Lamin Y. Darboe, MRCG at LSHTM for managing the study database. The authors would also like to thank the population of West Kiang District, The Gambia for participating in the study. The authors acknowledge Dr Branwen Henning for setting up the Keneba Biobank.

Author contributions

MWJ, AMP and CC conceptualized the study. MWJ, SC and CC designed the study. MWJ and AS conducted data collection and laboratory analysis; MWJ, AMP and CC performed the data analysis; MWJ wrote the manuscript. All authors revised and approved the final manuscript.

6.7. References

1. Lachowicz JI, Nurchi VM, Fanni D, Gerosa C, Peana M, Zoroddu MA. Nutritional Iron Deficiency: The Role of Oral Iron Supplementation. CMC. 2014 Jul 6;21(33):3775–84.
2. Camaschella C. Iron deficiency. Blood. 2019 Jan 3;133(1):30–9.
3. Kassebaum NJ. The Global Burden of Anemia. Hematology/Oncology Clinics of North America. 2016 Apr;30(2):247–308.
4. Finberg KE, Whittlesey RL, Fleming MD, Andrews NC. Down-regulation of Bmp/Smad signaling by Tmprss6 is required for maintenance of systemic iron homeostasis. Blood. 2010 May 6;115(18):3817–26.
5. Camaschella C, Nai A, Silvestri L. Iron metabolism and iron disorders revisited in the hepcidin era. Haematologica. 2020 Feb;105(2):260–72.
6. Silvestri L, Pagani A, Nai A, De Domenico I, Kaplan J, Camaschella C. The Serine Protease Matriptase-2 (TMPRSS6) Inhibits Hepcidin Activation by Cleaving Membrane Hemojuvelin. Cell Metabolism. 2008 Dec;8(6):502–11.
7. Finberg KE, Heeney MM, Campagna DR, Aydınok Y, Pearson HA, Hartman KR, et al. Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA). Nat Genet. 2008 May;40(5):569–71.
8. Gichohi-Wainaina WN, Towers GW, Swinkels DW, Zimmermann MB, Feskens EJ, Melse-Boonstra A. Inter-ethnic differences in genetic variants within the transmembrane protease, serine 6 (TMPRSS6) gene associated with iron status indicators: a systematic review with meta-analyses. Genes Nutr. 2015 Jan;10(1):442.

9. Nai A, Pagani A, Silvestri L, Campostrini N, Corbella M, Girelli D, et al. TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. Blood. 2011 Oct 20;118(16):4459–62.
10. Tanaka T, Roy CN, Yao W, Matteini A, Semba RD, Arking D, et al. A genome-wide association analysis of serum iron concentrations. Blood. 2010 Jan 7;115(1):94–6.
11. Delbini P, Vaja V, Graziadei G, Duca L, Nava I, Refaldi C, et al. Genetic variability of TMPRSS6 and its association with iron deficiency anaemia: Short Report. British Journal of Haematology. 2010 Nov;151(3):281–4.
12. Benyamin B, Ferreira MAR, Willemsen G, Gordon S, Middelberg RPS, McEvoy BP, et al. Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. Nat Genet. 2009 Nov;41(11):1173–5.
13. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68–74.
14. Nai A, Pagani A, Silvestri L, Campostrini N, Corbella M, Girelli D, et al. TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. Blood. 2011 Oct 20;118(16):4459–62.
15. Lee P. Role of Matriptase-2 (TMPRSS6) in Iron Metabolism. Acta Haematol. 2009;122(2–3):87–96.
16. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet. 2011 Oct;12(10):683–91.
17. Gan W, Guan Y, Wu Q, An P, Zhu J, Lu L, et al. Association of TMPRSS6 polymorphisms with ferritin, hemoglobin, and type 2 diabetes risk in a Chinese

Han population. The American Journal of Clinical Nutrition. 2012 Mar 1;95(3):626–32.

18. Jallow MW, Cerami C, Clark TG, Prentice AM, Campino S. Differences in the frequency of genetic variants associated with iron imbalance among global populations. PLoS ONE. 2020 Jul 1;15(7):e0235141.

19. Jallow MW, Campino S, Prentice AM, Cerami C. A recall-by-genotype study on polymorphisms in the TMPRSS6 gene and oral iron absorption: a study protocol. F1000Research. 2019 May 21;8(May):701.

20. Hennig BJ, Unger SA, Dondeh BL, Hassan J, Hawkesworth S, Jarjou L, et al. Cohort profile: The Kiang West Longitudinal Population Study (KWLPS)-a platform for integrated research and health care provision in rural Gambia. International Journal of Epidemiology. 2017;46(2):1–12.

21. Hwang S-I, Lee Y-Y, Park J-O, Norton HJ, Clemens E, Schrum LW, et al. Effects of a single dose of oral iron on hepcidin concentrations in human urine and serum analyzed by a robust LC-MS/MS method. Clinica Chimica Acta. 2011 Nov;412(23–24):2241–7.

22. Suchdev PS, Williams AM, Mei Z, Flores-Ayala R, Pasricha S-R, Rogers LM, et al. Assessment of iron status in settings of inflammation: challenges and potential approaches. The American Journal of Clinical Nutrition. 2017 Dec;106(Supplement 6):1626S-1633S.

23. Shattnawi KK, Alomari MA, Al-Sheyab N, Bani Salameh A. The relationship between plasma ferritin levels and body mass index among adolescents. Scientific Reports. 2018 Dec 17;8(1):15307.

24. R Core Team. A Language and Environment for Statistical Computing. Vol. 3.5.1. Vienna, Austria: R Foundation for Statistical Computing; 2018.

25. Blanco-Rojo R, Baeza-Richer C, López-Parra AM, Pérez-Granados AM, Brichs A, Bertoncini S, et al. Four variants in transferrin and HFE genes as potential markers of iron deficiency anaemia risk: an association study in menstruating women. Nutrition & metabolism. 2011 Oct 6;8:69.

26. Cheng HL, Hancock DP, Rooney KB, Steinbeck KS, Griffin HJ, O'Connor HT. A candidate gene approach for identifying differential iron responses in young overweight women to an energy-restricted haem iron-rich diet. European Journal of Clinical Nutrition. 2014 Nov 7;68(11):1250–2.

27. Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, et al. Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. Nature Genetics. 2009 Nov 11;41(11):1170–2.

28. Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. Nature communications. 2014 Oct 29;5(2):4926.

29. Kullo IJ, Ding K, Jouni H, Smith CY, Chute CG. A Genome-Wide Association Study of Red Blood Cell Traits Using the Electronic Medical Record. Dubé M-P, editor. PLoS ONE. 2010 Sep 28;5(9):e13011.

30. Gichohi-Wainaina WN, Melse-Boonstra A, Swinkels DW, Zimmermann MB, Feskens EJ, Towers GW. Common Variants and Haplotypes in the TF, TNF-α, and TMPRSS6 Genes Are Associated with Iron Status in a Female Black South African Population. The Journal of Nutrition. 2015 May 1;145(5):945–53.

31. Lone NM, Shah SHS, Farooq M, Arif M, Younis S, Riaz S. Role of TMPRSS6 rs855791 (T > C) polymorphism in reproductive age women with iron deficiency anemia from Lahore, Pakistan. Saudi J Biol Sci. 2021 Jan;28(1):748–53.

32. Nalado AM, Dickens C, Dix-Peek T, Mahlangu JN, Olorunfemi G, Paget G, et al. TMPRSS6 rs855791 polymorphism and susceptibility to iron deficiency anaemia in non-dialysis chronic kidney disease patients in South Africa. Int J Mol Epidemiol Genet. 2019;10(1):1–9.

33. Buerkli S, Pei S-N, Hsiao S-C, Lee C-T, Zeder C, Zimmermann MB, et al. The TMPRSS6 variant (SNP rs855791) affects iron metabolism and oral iron absorption – a stable iron isotope study in Taiwanese women. haematol. 2020 Oct 5;0–0.

34. Ganz T. Systemic Iron Homeostasis. Physiological Reviews. 2013 Oct;93(4):1721–41.

35. Girelli D, Nemeth E, Swinkels DW. Hepcidin in the diagnosis of iron disorders. Blood. 2016 Jun 9;127(23):2809–13.

36. Pagani A, Nai A, Silvestri L, Camaschella C. Hepcidin and Anemia: A Tight Relationship. Front Physiol. 2019 Oct 9;10:1294.

37. Corbin LJ, Tan VY, Hughes DA, Wade KH, Paul DS, Tansey KE, et al. Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference. Nature Communications. 2018 Dec 19;9(1):711.

38. Maharaj S, Seegobin K, Shaikh M. Fasting vs. postprandial evaluation of iron deficiency anemia in cancer. JCO. 2018 May 20;36(15_suppl):e18876–e18876.

6.8. Tables

Table 1. The description of the configuration of the genotype combinations that formed the bases of the participant selection.

Genotype Combination	**rs2235321** major/ minor allele	**rs855791** major/ minor allele	**rs4820268** major/ minor allele	**Individuals available for selection**	**Genotype Frequency**[1]
	G/A	**G**/A	**A**/G		
GG/GG/AA	**G/G**	**G/G**	**A/A**	117	0.066
AA/**GG**/AA	A/A	**G/G**	**A/A**	336	0.190
A**G**/**GG**/G**A**	A/**G**	**G/G**	G/**A**	391	0.229
GG/GG/G**A**	**G/G**	**G/G**	G/**A**	211	0.129
GG/GG/GG	**G/G**	**G/G**	G/G	92	0.054
A**G**/A**G**/**AA**	A/**G**	A/**G**	**A/A**	60	0.044
A**G/GG**/**AA**	A/**G**	**G/G**	**A/A**	361	0.211
GG/A**G**/**AA**	**G/G**	A/**G**	**A/A**	67	0.035
GG/A**G**/G**A**	**G/G**	A/**G**	**G**/A	60	0.033

Bolded alleles indicate major alleles. The genotype combination GG/GG/AA consists of homozygotes for the major alleles at all the three SNPs. We used this group as the reference group.

[1]The frequency of each genotype combination in the population with genotype data in The Keneba Biobank (n=3116)

The bolded letters indicate the major alleles for each SNPs.

Table 2. The number of individuals enrolled into each genotype group.

Genotype group	n	Number of minor alleles
Reference group (GG/GG/AA)	39	0
rs2235321, A/A (AA/GG/AA)	35	2
rs2235321, A/G (AG/GG/AA)	21	1
rs855791, A/G (GG/AG/AA)	28	1
rs4820268, G/A (GG/GG/GA)	28	1
rs4820268, G/G (GG/GG/GG)	29	2
Double heterozygote (AG/AG/AA: rs2235231 A/G & rs855791 A/G)	13	2
Double heterozygote (AG/GG/GA: rs2235321 A/G & rs4820268 G/A)	38	2
Double heterozygote (GG/AG/GA: rs855791 A/G & rs4820268 G/A)	20	2
Total number of study participants (N)	251	

The details of the genotype group configuration are presented is **Table 1**

.

Table 3. Baseline characteristics of the study population

Variable	All (n=251)	F (n=191)	M (n=60)	*p*-value (F vs M)
Age, yrs[1]	29.0 (18.0, 50.0)	33.0 (18.0, 50.0)	22.0 (18.0, 40.0)	<0.001
Plasma iron (µmol/L)[1]	13.5 (0.4, 57.1)	12.3 (0.4, 57.1)	15.6 (4.4, 36.2)	0.012
Hepcidin (ng/mL)[1]	2.89 (0.05, 71.70)	2.46 (0.05, 34.60)	3.86 (0.09, 71.71)	0.009
TSAT (%)[1]	21.6 (0.6, 100.0)	20.8 (0.6, 100)	26.0 (6.0, 57.8)	0.001
Transferrin (g/L)[2]	2.82 (0.58)	2.91 (0.55)	2.55 (0.61)	<0.001
UIBC (µmol/L)[1]	46.9 (21.1, 105.1)	48.2 (21.1, 105.1)	41.7 (22.2, 86.2)	<0.001
TIBC (µmol/L)[1]	61.4 (10.3, 112.2)	62.2 (10.3, 112.2)	57.0 (35.9, 94.7)	0.002
Ferritin (µg/L)[1]	31.0 (0.0, 237.7)	25.5 (0.0, 237.7)	50.0 (7.8, 160.4)	<0.001
sTfR (mg/L)[1]	4.00 (1.90, 11.32)	4.11 (1.90 11.32)	3.54 (2.01 7.62)	0.011
CRP (mg/L)[1]	0.80 (0.03, 26.95)	0.91 (0.03, 26.95)	0.63 (0.05, 13.40)	0.175
Hb (g/dL)[2]	12.3 (1.5)	12.0 (1.3)	13.3 (1.6)	<0.001
RBC (x10^12)[2]	4.4 (0.6)	4.3 (0.5)	4.8 (0.6)	<0.001
MCV (fL)[2]	81.1 (6.1)	81.1 (6.0)	80.0 (6.2)	0.882
Haematocrit (%)[2]	35.9 (4.5)	35.0 (3.8)	38.8 (5.0)	<0.001
RDW (%)[2]	12.2 (1.2)	12.2 (1.2)	12.3 (1.0)	0.615
MCH (pg)[2]	27.9 (2.5)	28.0 (2.5)	27.8 (2.5)	0.675
MCHC (g/dl)[2]	34.4 (1.2)	34.5 (1.2)	34.3 (1.1)	0.424
BMI (Kg/m2)[1]	21.3 (14.4, 39.1)	22.0 (14.4, 39.1)	19.5 (16.3, 25.8)	<0.001
Sickle cell trait (AS/AA)[3]	21/251	17/191	4/60	0.742
G6PD deficiency (carrier/non-carrier)[3,4]	4/177	0/133	4/44	0.003

[1]Skewed data listed as medians (ranges)
[2]Normally distributed data presented as means (SD)
[3]Categorical data presented as proportions
Student *t*-test was used to determine the differences between parametric data, and Wilcoxon test was used for non-parametric data. Chi-square was used to test the differences between categorical data.
[4]Individuals not tested for G6PD deficiency (n=74).

TSAT, transferrin saturation; UIBC, unsaturated iron binding capacity; TIBC, total iron binding capacity; sTfR, soluble transferrin receptor; CRP, C-reactive protein; Hb, haemoglobin; RBC, red blood cell number; MCV, mean corpuscular volume; RDW, red cell distribution width; MCH, mean corpuscular haemoglobin; MCHC, mean corpuscular haemoglobin concentration; BMI, body mass index; G6PD, glucose-6-phosphase dehydrogenase

TSAT, transferrin saturation; UIBC, unsaturated iron binding capacity; TIBC, total iron binding capacity; sTfR, soluble transferrin receptor; CRP, C-reactive protein; Hb, haemoglobin; RBC,

6.9. Figure Legends

Figure 1. A flow chart illustrating how the study participants were selected into nine groups based on the three candidate *TMPRSS6* SNPs. The reference group consisted of individuals who were homozygous for the major alleles at all the three SNPs. The configuration of the genotype combinations is presented in **Supplemental Table 1**.

Figure 2. Plasma iron concentrations in each genotype group before and after iron ingestion. Plasma iron significantly increased in all the genotype groups after the iron dose. The reference is the genotype group with individuals that carries two major allele at all the three SNPs. The configuration of the genotype groups is presented in **Table 1**.

The horizontal lines showing the P values indicates the difference between the two-extreme bar.

Figure 3. The differences between genotypes of individual SNPs (rs2235321, rs85579 and rs4820268) (**A**), and between the double heterozygotes and the reference group (**B**), on iron concentration before and after iron ingestion.

The reference group consisted of individuals who are homozygotes for the major allele at all the three SNPs (rs2235321 GG, rs855791 GG and rs4820268 AA), see **Tables 1** and **2**. There was no significant differences between the genotype groups in iron concentration both before and after the iron dose.

Figure 4. Hepcidin concentrations before and after the iron dose, within the genotypes of individual SNPs (**A**) and in the double heterozygotes and the reference group (**B**). Hepcidin increased in all the genotype groups except in the genotype group AG/AG/AA (double heterozygotes at rs2235321 and rs855791).

The horizontal lines showing the P values indicates the difference between the two extreme bar.

Figure 5. Differences in hepcidin concentrations between genotypes of individual SNPs (**A**), and between double heterozygotes and the reference group (**B**), before and after iron ingestion.

The horizontal lines showing the P values in Figure A indicates the differences between the two extreme bar for each SNP.

For each SNP, carriage of the homozygous major alleles is associated with elevated hepcidin concentrations. The genotype group AG/AG/AA had the lowest hepcidin concentration both before and after the iron dose (**B**). The reference group consist of individuals without any minor allele from the three SNPs (rs2235321 GG, rs855791 GG and rs4820268 AA), see **Tables 1** and **2**. The group AG/AG/AA (consisting of heterozygous at rs2235321 and rs855791, and homozygous major allele at rs4820268).

6.10. Figures

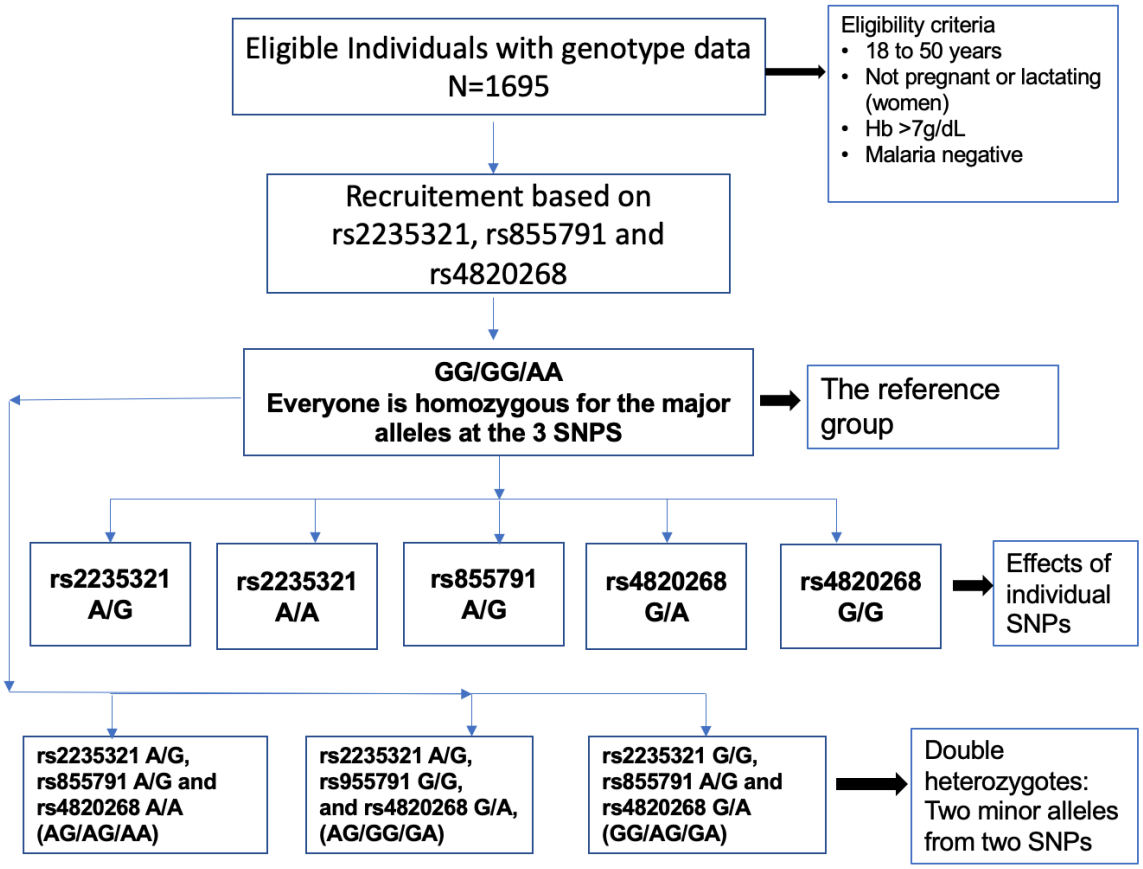


Figure 1

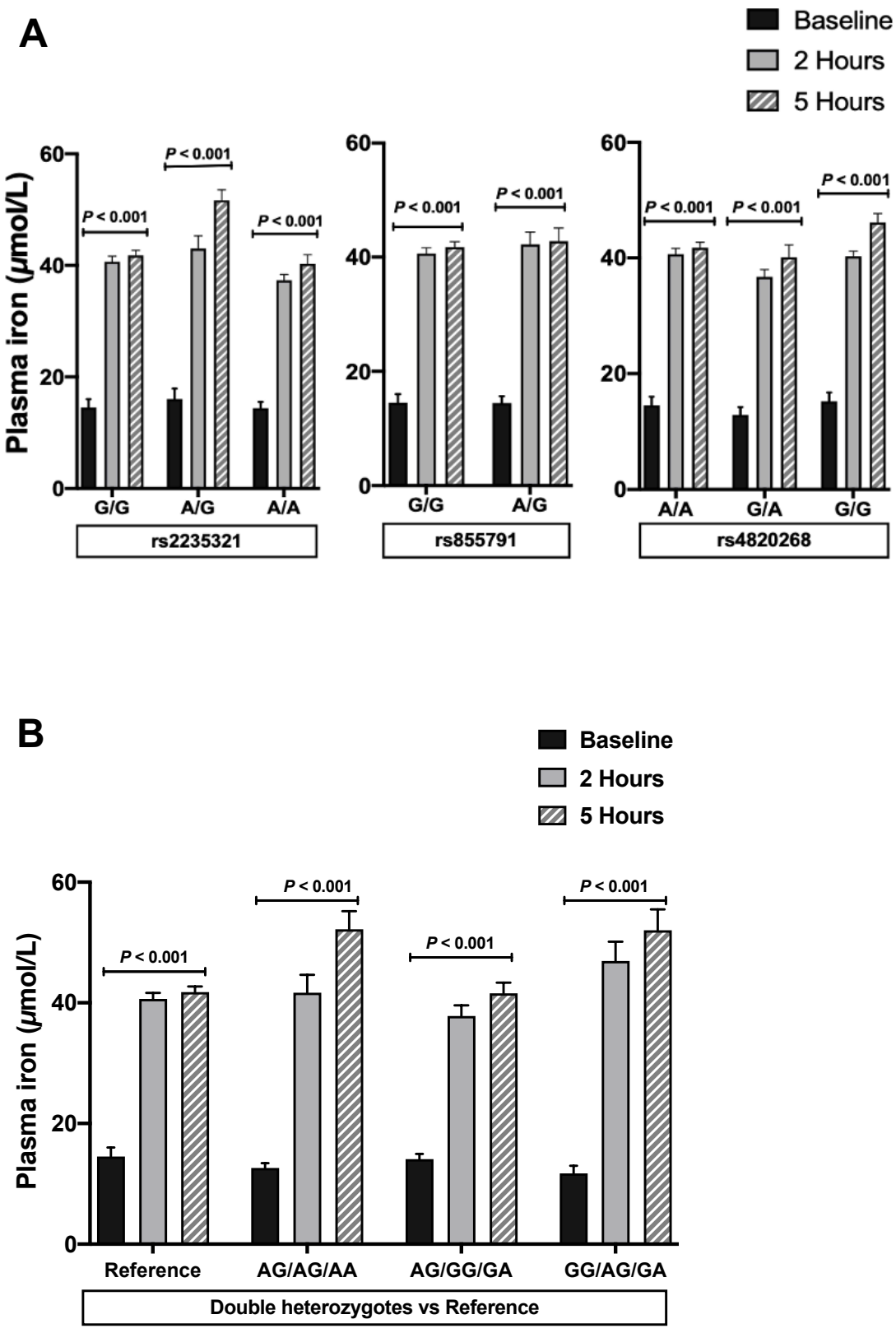
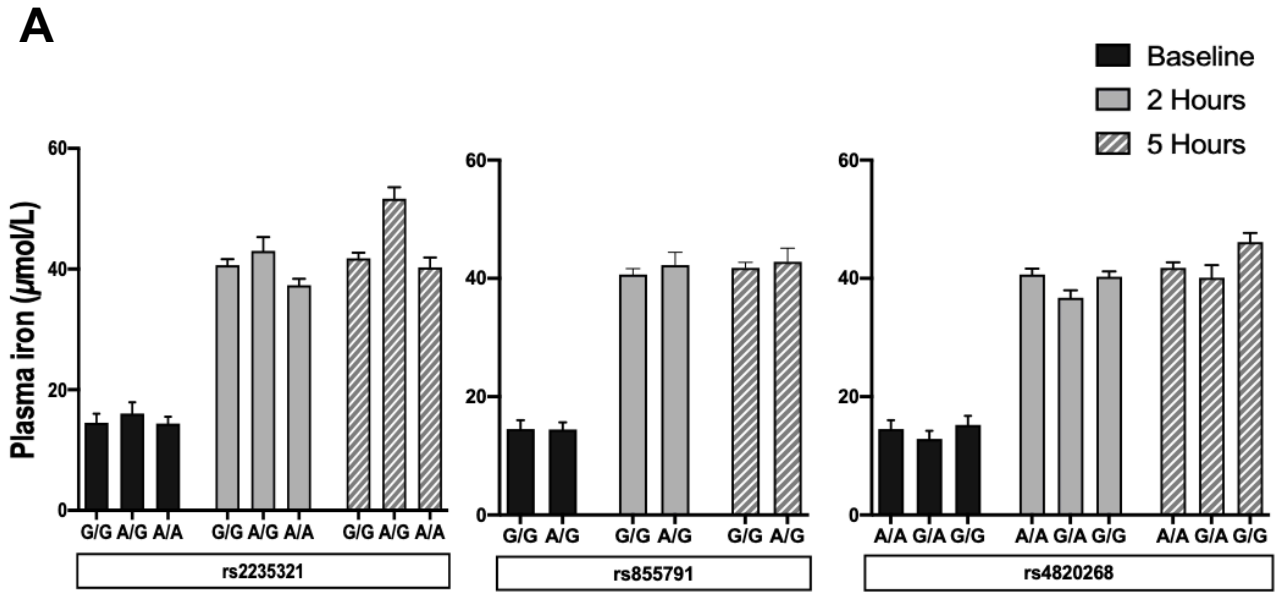

A
Baseline
2 Hours
5 Hours
Plasma iron (μmol/L)
60
40
20
0
P < 0.001
G/G
A/G
A/A
rs2235321
rs855791
A/A
G/A
G/G
rs4820268
B
Reference
AG/AG/AA
AG/GG/GA
GG/AG/GA
Double heterozygotes vs Reference

Figure 2


A
Baseline
2 Hours
5 Hours
Plasma iron (µmol/L)
60
40
20
0
G/G A/G A/A
G/G A/G A/A
G/G A/G A/A
rs2235321
G/G A/G
G/G A/G
G/G A/G
rs855791
A/A G/A G/G
A/A G/A G/G
A/A G/A G/G
rs4820268

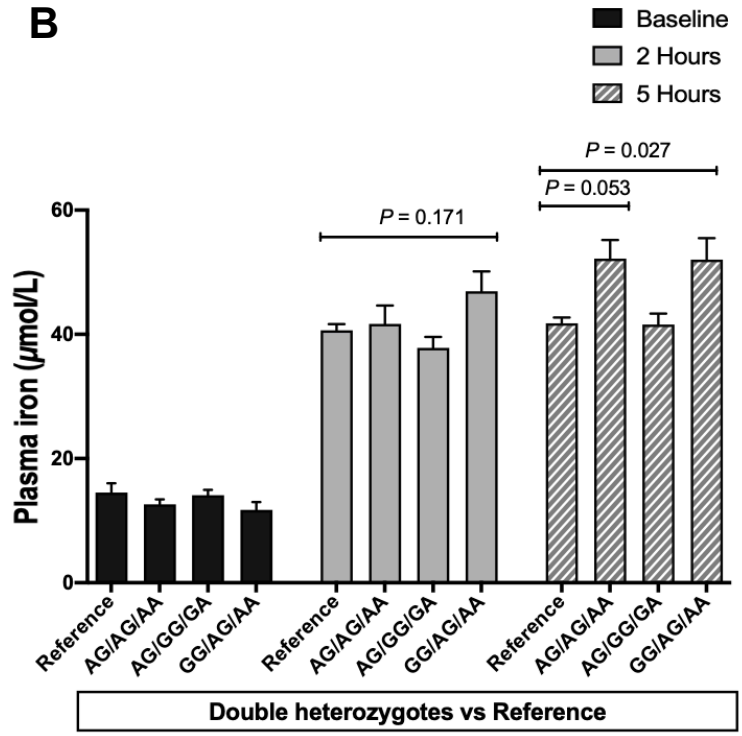

B
Baseline
2 Hours
5 Hours
P = 0.027
P = 0.053
P = 0.171
Plasma iron (µmol/L)
60
40
20
0
Reference
AG/AG/AA
AG/GG/GA
GG/AG/AA
Reference
AG/AG/AA
AG/GG/GA
GG/AG/AA
Reference
AG/AG/AA
AG/GG/GA
GG/AG/AA
Double heterozygotes vs Reference

Figure 3

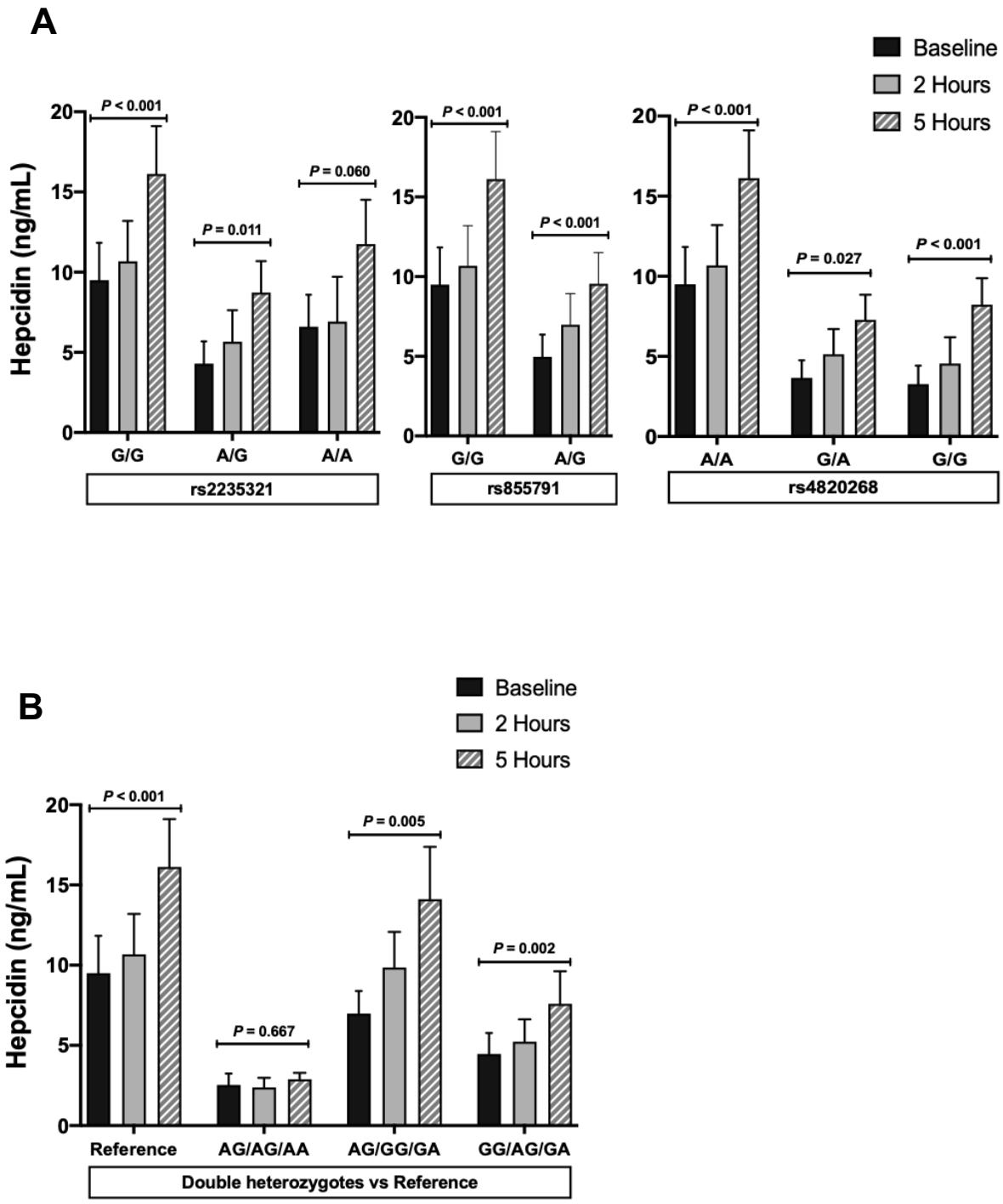
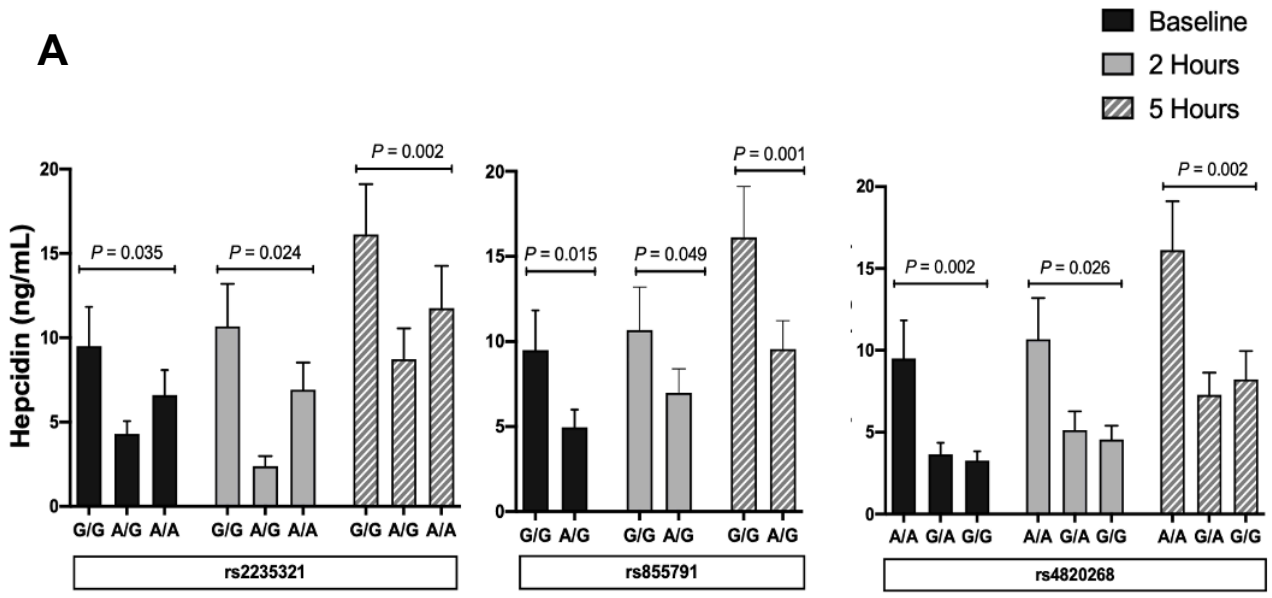

A
Baseline
2 Hours
5 Hours
Hepcidin (ng/mL)
P < 0.001
P = 0.011
P = 0.060
P < 0.001
P < 0.001
P < 0.001
P = 0.027
P < 0.001
G/G
A/G
A/A
G/G
A/G
A/A
G/A
G/G
rs2235321
rs855791
rs4820268
B
Baseline
2 Hours
5 Hours
Hepcidin (ng/mL)
P < 0.001
P = 0.667
P = 0.005
P = 0.002
Reference
AG/AG/AA
AG/GG/GA
GG/AG/GA
Double heterozygotes vs Reference

Figure 4


A
Baseline
2 Hours
5 Hours
Hepcidin (ng/mL)
P = 0.035
P = 0.024
P = 0.002
P = 0.015
P = 0.049
P = 0.001
P = 0.002
P = 0.026
P = 0.002
G/G A/G A/A
rs2235321
rs855791
A/A G/A G/G
rs4820268

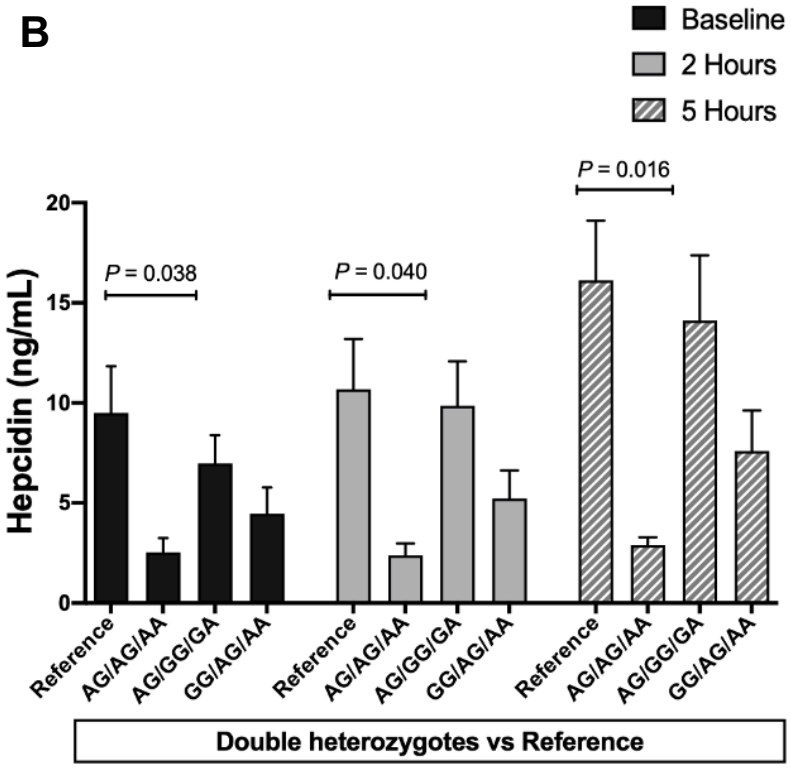

B
Baseline
2 Hours
5 Hours
Hepcidin (ng/mL)
P = 0.038
P = 0.040
P = 0.016
Reference
AG/AG/AA
AG/GG/GA
GG/AG/AA
Double heterozygotes vs Reference

Figure 5

6.11. Supplemental information

Common variants in the transmembrane protease serine 6 (*TMPRSS6*) gene alter hepcidin but not oral iron absorption in healthy Gambian adults: a recall-by-genotype study

Authors:

Momodou W. Jallow, Susana Campino, Alasana Saidykhan, Andrew M. Prentice and Carla Cerami

Supplemental Table 1. The effects of individuals SNPs on iron markers, before and after iron ingestion

Genotype	Baseline				2 Hours				5 Hours			
	Serum iron (umol/L)											
	Mean	**SE**	**Beta**	***p*-value**	**Mean**	**SE**	**Beta**	***p*-value**	**Mean**	**SE**	**Beta**	***p*-value**
Reference	14.53	1.32	**Reference**		40.66	2.69	**Reference**		41.78	2.67	**Reference**	
rs855791 (A/**G**)	14.46	2.04	-0.07	0.972	42.26	4.05	1.60	0.755	42.82	4.05	1.037	0.939
rs2235321 (A/**G**)	16.04	2.25	1.51	0.473	43.01	4.22	2.35	0.810	51.68	4.35	9.90	0.048
rs2235321 (A/A)	14.38	1.94	-0.15	0.738	37.33	3.64	-3.34	0.620	40.29	3.77	-1.49	0.607
rs4820268 (G/**A)**	12.86	2.09	-1.67	0.503	36.71	3.63	-3.95	0.490	40.11	4.14	-1.67	0.506
rs4820268 (G/G)	15.24	2.07	0.72	0.589	40.27	3.60	-0.39	0.984	46.13	4.14	4.35	0.450

Effects of double heterozygotes on plasma iron before and after iron ingestion

Genotype	Baseline				2 Hours				5 Hours			
	Serum iron (umol/L)											
	Mean	**SE**	**Beta**	***p*-value**	**Mean**	**SE**	**Beta**	***p*-value**	**Mean**	**SE**	**Beta**	***p*-value**
Reference	14.53	1.12	**Reference**		40.66	2.67	**Reference**		41.78	2.67	**Reference**	
A**G**/A**G**/**AA**	12.62	2.30	-1.91	0.409	41.69	5.30	1.03	0.846	52.18	5.33	10.41	0.054
A**G**/**GG**/G**A**	14.09	1.59	-0.44	0.782	37.81	3.78	-2.86	0.452	41.57	3.79	-0.21	0.955
GG/A**G**/G**A**	11.73	1.92	-2.80	0.148	46.94	4.55	6.27	0.171	52.03	4.58	10.25	0.027

Supplemental Table 1 Continued

Genotypes	Baseline				2 Hours				5 Hours			
	Transferrin saturation (%)											
	Mean	**SE**	**Beta**	***p*-value**	**Mean**	**SE**	**Beta**	***p*-value**	**Mean**	**SE**	**Beta**	***p*-value**
Reference	2.90	0.10	**Reference**		4.06	0.07	**Reference**		4.16	0.06	**Reference**	
rs2235321 (A/**G**)	3.11	0.17	0.20	0.224	4.09	0.12	0.04	0.759	4.35	0.11	0.19	0.084
rs2235321 (A/A)	3.10	0.14	0.20	0.173	4.08	0.10	0.02	0.814	4.17	0.09	0.01	0.957
rs855791 (A/**G**)	3.13	0.16	0.22	0.160	4.14	0.10	0.08	0.442	4.16	0.10	0.00	0.961
rs4820268 (G/**A**)	2.88	0.16	-0.02	0.884	4.17	0.11	-0.05	0.884	4.06	0.11	-0.11	0.307
rs4820268 (G/G)	3.13	0.15	0.23	0.143	4.16	0.11	0.14	0.629	4.32	0.11	0.15	0.153
	Transferrin (g/L)											
Reference	2.86	0.08	**Reference**		2.79	0.09	**Reference**		2.64	0.08	**Reference**	
rs2235321 (A/**G**)	2.83	0.14	-0.03	0.842	2.90	0.15	0.11	0.456	2.82	0.14	0.18	0.206
rs2235321 (A/A)	2.68	0.12	-0.18	0.133	2.81	0.13	0.01	0.924	2.73	0.12	0.09	0.468
rs855791 (A/**G**)	2.81	0.14	-0.04	0.752	2.80	0.15	0.01	0.971	2.76	0.14	0.12	0.414
rs4820268 (G/**A**)	3.08	0.14	0.22	0.118	3.12	0.16	0.33	0.040	3.11	0.15	0.46	0.227
rs4820268 (G/G)	2.80	0.14	-0.06	0.652	2.87	0.16	0.07	0.645	2.92	0.15	0.28	0.102

Supplemental Table 1 Continued

Genotypes	Baseline				2 Hours				5 Hours			
	Total iron binding capacity (μmol/L)											
	Mean	**SE**	**Beta**	***p*-value**	**Mean**	**SE**	**Beta**	***p*-value**	**Mean**	**SE**	**Beta**	***p*-value**
Reference	68.29	1.81	**Reference**		65.16	1.88	**Reference**		60.93	1.62	**Reference**	
rs2235321 (A/**G**)	61.63	3.06	-6.66	**0.032**	64.15	3.15	-1.01	0.750	63.54	2.74	2.61	0.344
rs2235321 (A/A)	58.78	2.63	-9.51	**0.000**	60.21	2.71	-4.94	0.072	57.81	2.37	-3.12	0.192
rs855791 (A/**G**)	60.79	3.50	-7.50	**0.036**	62.36	3.07	-2.80	0.365	60.82	2.53	-0.12	0.963
rs4820268 (G/**A**)	63.33	2.89	-4.96	**0.089**	64.43	3.15	-2.22	0.482	63.82	2.81	1.97	0.483
rs4820268 (G/G)	58.76	2.86	-9.54	**0.001**	61.92	3.12	-8.18	**0.010**	59.57	2.81	-3.20	0.258
	Unsaturated iron binding capacity (μmol/L)											
Reference group	53.76	2.27	**Reference**		24.49	2.52	**Reference**		19.15	2.31	**Reference**	
rs2235321 (A/**G**)	45.59	3.83	-8.17	0.036	21.14	4.22	-3.35	0.429	11.86	3.91	-7.29	0.065
rs2235321 (A/A)	44.40	3.30	-9.36	**0.006**	22.89	3.64	-1.60	0.660	17.52	3.39	-1.64	0.630
rs855791 (A/**G**)	48.05	3.91	-5.71	0.15	20.11	3.70	-4.38	0.240	18.32	3.49	-0.84	0.811
rs4820268 (G/**A**)	50.47	3.64	-3.29	0.368	26.22	3.53	1.73	0.626	22.79	3.46	3.64	0.296
rs4820268 (G/G)	43.51	3.61	-10.25	**0.005**	16.71	3.50	-7.79	0.028	11.99	3.42	-7.16	0.039

For transferrin saturation, statistical analysis was done on log-transformed data, but transferrin TIBC and UIBC values were not log-transformed as these values were normally distributed. For each of the SNPs, the bolded letters are the major allele. The reference group consists of individuals with homozygotes for the major alleles at all the three SNPs. The participants were selected in such a way that the reference genotype for each SNP do not have individual carrying the minor alleles at the other two SNPs.

SE, standard error; TIBC, total iron binding capacity; TSAT, transferrin saturation; UIBC, unsaturated iron binding capacity.

Supplemental Table 2. The effects of individuals SNPS (rs855791, rs2235321 and rs4820268) on haematology traits

rs2235321 (MAF=0.43)					
Trait	**Genotype**	**Mean**	**Std. Error**	**Beta**	***P*-value**
Hb (g/dL)	**G/G**	11.6	0.33	**Reference**	
	A/**G**	11.9	0.46	0.27	0.551
	A/A	12.1	0.38	0.43	0.262
RBC (x10^12)	**G/G**	4.4	0.12	**Reference**	
	A/**G**	4.4	0.17	-0.05	0.749
	A/A	4.2	0.14	-0.26	0.072
MCV (fL)	**G/G**	77.8	1.54	**Reference**	
	A/**G**	79.5	2.13	1.63	0.445
	A/A	83.7	1.79	5.81	**0.002**
HCT (%)	**G/G**	34.4	1.03	**Reference**	
	A/**G**	34.8	1.42	0.34	0.814
	A/A	34.9	1.20	0.49	0.685
RDW (%)	**G/G**	13.1	0.29	**Reference**	
	A/**G**	12.7	0.40	-0.38	0.342
	A/A	12.2	0.33	-0.88	**0.010**
MCH (pg)	**G/G**	26.4	0.60	**Reference**	
	A/**G**	27.3	0.83	0.88	0.292
	A/A	29.0	0.70	2.55	**0.000**
MCHC (g/dL)	**G/G**	33.9	0.25	**Reference**	
	A/**G**	34.3	0.34	0.34	0.321
	A/A	34.6	0.29	0.69	**0.019**
rs855791 (MAF=0.07)					
Hb (g/dL)	**G/G**	11.9	0.44	**Reference**	
	A/**G**	12.0	0.46	0.07	0.882
RBC (x10^12)	**G/G**	4.5	0.14	**Reference**	
	A/**G**	4.5	0.17	-0.04	0.832
MCV (fL)	**G/G**	78.8	1.56	**Reference**	
	A/**G**	81.0	1.97	2.29	0.251
HCT (%)	**G/G**	35.2	1.18	**Reference**	
	A/**G**	35.9	1.49	0.70	0.641
RDW (%)	**G/G**	12.8	0.28	**Reference**	
	A/**G**	12.5	0.36	-0.36	0.311
MCH (pg)	**G/G**	26.7	0.61	**Reference**	
	A/**G**	27.1	0.77	0.39	0.617
MCHC (g/dL)	**G/G**	33.8	0.30	**Reference**	
	A/**G**	33.4	0.38	-0.48	0.207

Supplemental Table 2. Continued

rs4820268 (MAF=0.27)					
Trait	**Genotype**	**Mean**	**SE**	**Beta**	***P*-value**
Hb (g/dL)	**A/A**	12.09	0.23	**Reference**	
	G/**A**	11.69	0.32	-0.40	0.213
	G/G	11.82	0.31	0.13	0.687
RBC (x10^12)	**A/A**	4.46	0.09	**Reference**	
	G/**A**	4.27	0.12	-0.19	0.125
	G/G	4.38	0.12	-0.07	0.539
MCV (fL)	**A/A**	81.37	0.98	**Reference**	
	G/**A**	78.67	1.50	-2.70	0.076
	G/G	79.80	1.49	-1.57	0.294
HCT (%)	**A/A**	35.94	0.71	**Reference**	
	G/**A**	33.38	1.00	-2.56	0.012
	G/G	34.66	0.98	-1.28	0.192
RDW (%)	**A/A**	12.29	0.21	**Reference**	
	G/**A**	12.54	0.32	0.25	0.449
	G/G	12.18	0.32	-0.12	0.709
MCH (pg)	**A/A**	27.46	0.39	**Reference**	
	G/**A**	27.57	0.60	0.11	0.861
	G/G	28.15	0.59	0.69	0.249
MCHC (g/dL)	**A/A**	33.73	0.18	**Reference**	
	G/**A**	34.99	0.28	1.26	0.000
	G/G	35.25	0.28	1.53	0.000

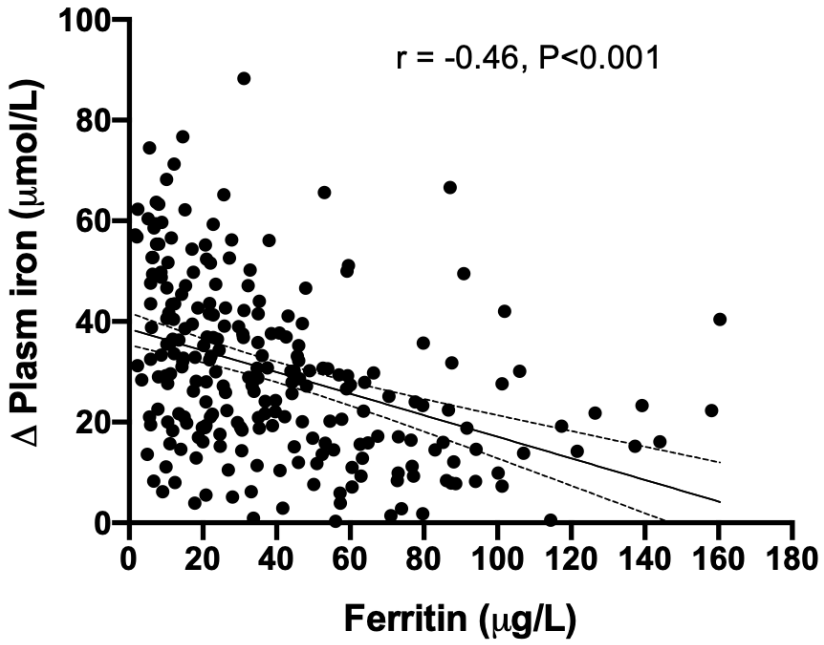


Figure S1. The relationship between baseline ferritin and delta plasma iron. Delta plasma iron was calculated by subtracting the baseline plasma iron from the 5 hours plasma iron post iron ingestion.

6.12. A summary of the pilot study

In order to decide the optimal time points for the recall-by-genotype study, we conducted a pilot study. In this pilot stud, we selected participants from five genotype groups based on *TMPRSS6* rs22235321 and rs4820268, while ensuring that no one carries the variant allele at rs855791, see **Table 1**.

We included individuals if they were 18 years and above, although there is no maximum age limit, individual considered too old or frail were excluded. Individuals were excluded if they reported being unwell or when they were found to be unwell upon examination by a study nurse. Also, pregnant or lactating women, and individuals with severe anaemia (Hb <7g/dl) where excluded.

Iron supplementation and sample collection

Each individual received 400mg ferrous sulfate (130 mg of elemental iron) orally (after an overnight fast) on Day 1, and then daily supplementation with 200mg of ferrous sulfate (65 mg of elemental iron) from day 2 for 14 days. At day one, each individual was bled at baseline following an overnight fast (before receiving the supplement) and at 1h, 2h, 5h, 24h. Thereafter, a blood sample was taken at (day 15), after iron completing the daily iron dose.

Laboratory measurements

FBC was analysed using a 3-part haematology analyser Medonic M-series (Boule Medical, Sweden). Malaria rapid diagnostic test (SD BioLine Malaria Antigen Pf, Standard Diagnostics Inc. Republic of Korea), sickle (Sodium metabisulphide method) and quantitative assessment of G6PD enzyme in RBCs (G6PD Hb+ R&D Diagnostics) were done on the same day.

Iron biomarkers (serum iron, ferritin, serum transferrin, unsaturated iron binding capacity [UIBC]) and CRP were measured on frozen plasma samples by Cobas Integra 400Plus biochemistry analyser (Roche Diagnostics). Transferrin saturation (TSAT) and total iron binding capacity (TIBC) were calculated from UIBC and serum iron; TIBC= serum iron +UIBC; TSAT= (Serum iron/TIBC)*100. Hepcidin was measured by Bachem competitive ELISA (enzyme-linked immunosorbent assay) kit.

Data analysis

Data was analysed using the R statistical software version 3.2.3 [26]. Initially, a linear model with each biomarker as the outcome, genotype as response variable and age and sex as covariates was fitted. A more complex linear model was fitted to determine

the changes in biomarker across time points (using the baseline as a reference time group) and genotype as response variables. Where there was evidence of significant time or genotype effects, interactions between them was assessed using a log-likelihood ratio test.

Results

A total of 44 individuals were enrolled in 5 genotype groups (**Table 1**) using the *TMPRSS6* rs2235321 and rs4820248 as the basis for creating the genotype groups. From the 44 participants enrolled, 65.9% were females. The details of participant characteristics are presented in **Table 4.**

Table 1. Description of genotype combinations recruited for the pilot study, N=44

	rs2235321	**rs4820268**	**N (%)**
Major allele /minor allele	**G**/A	A/G	
AA/**AA**	AA	**AA**	8 (18.2)
A**G**/G**A**	A**G**	G**A**	10(22.7)
GG/**AA**	**GG**	**AA**	11(25.0)
GG/G**A**	**GG**	G**A**	3(6.8)
GG/GG	**GG**	GG	12(27.3)

Everyone in these genotype combinations are GG (homozygous for the major allele) at rs855791.
AA/AA, Homozygous for the variant of rs2235321, wildtype at rs4820268; AG/GA, Double heterozygotes, for both rs2235321 and rs4820268; GG/AA, Wildtype for rs2235321 and rs4820268 this is the control group; GG/GA, Heterozygous at rs4820268, wildtype at rs2235321; Homozygous for the variant allele of rs4820268, wildtype at rs2235321.

Table 2. Population characteristics of individuals enrolled in the pilot study

Variable	Male	Female	Total
N (%)	15 (34.1)	29 (65.9)	44
Age (yrs), mean (SD)	24.8 (5.5)	29.7 (8.5)	-
Sickle (n=44)			
AA (n=41, 93.2%))	14	27	41
AS (n=3, 6.8%)	1	2	3
G6PD (n=41)			
Non-deficient (n=38, 92.6%)	11	27	38
Deficient (n=3, 7.4%)	3	0	3

G6PD, glucose-6-phosphate dehydrogenase; AA and AS, sickle Hb wild-types AS heterozygotes respectively.

Assessing the change in iron biomarkers over time

Each iron biomarker was tested over six time-points [baseline and after iron supplementation [(1hr, 2hr, 5hr, 24hr and Day 15)]. The significant differences between time points in relation to the baseline occurred at 1hr, 2hr and 5hrs, in TSAT, hepcidin, serum iron and UIBC, see **Figure 1**. No significant changes in relation to the baseline were observed at 24hr and Day 15. Furthermore, no significant changes occurred in ferritin, transferrin, TIBC and CRP.

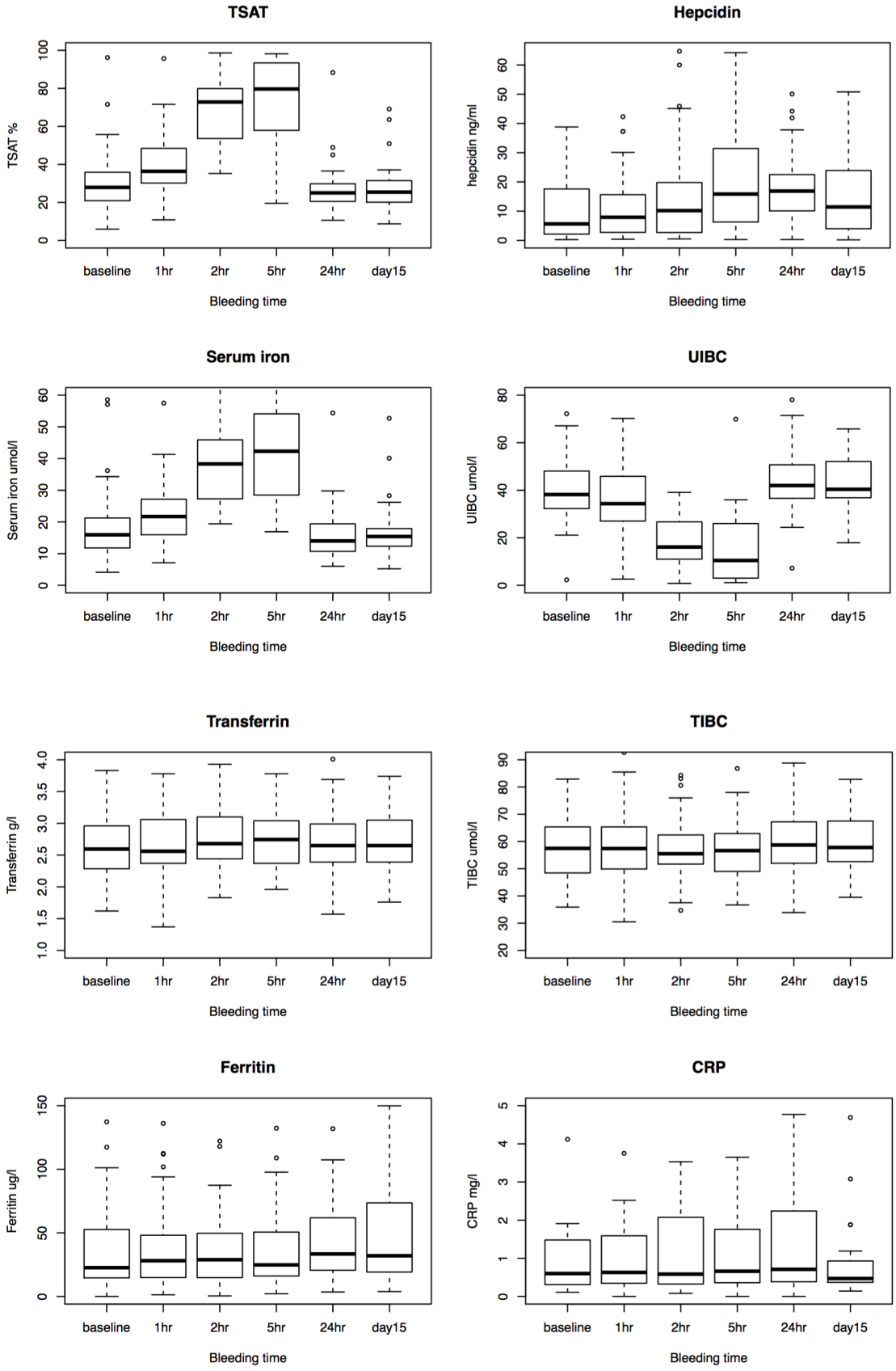


Figure 1. Assessment of important time-points for inclusion in the main study, N=44. Significant differences in relation to the baseline were observed in TSAT (1hr, p=0.007; 2hr, p<0.0001; 5hr, p<0.001), hepcidin (5hr, p=0.0032), serum iron (1hr, p=0.0325; 2hr, p<0.001; 5hr, p<0.001), UIBC (2hr, p<0.001; 5hr, p<0.001).

Optimisation of follow-up bleeds

Based on the response data from the pilot study, we selected time points 2 and 5 hours for post iron ingestion sampling. The 24 hours and Day 15 time points were dropped, as they do not provide additional information, and substantial change in iron status occurred within the first day, **see Figure 2**. This will reduce the burden on study participants, by shortening the study duration, while enabling us to achieve the aim of the study. Also, the 1hour bleeding was dropped as it is not expected to additional information.

Therefore, the selection of the time points for the main study (baseline, 2 and 5 hours) was based on the information derived from the pilot study as described above.

Chapter 7:

General discussion and conclusions

7.1. General discussion

This PhD thesis aimed to investigate whether common genetic variants within the hepcidin and iron regulatory genes, previously identified in non-African population, predispose healthy Gambians to anaemia and/or modulate response to oral iron supplementation. The work presented here was aimed at contributing to the understanding of genetic determinants of anaemia and oral iron absorption in Africans. In this chapter, I discuss the main findings of the studies constituting the thesis. Also, I highlight the limitations and presented suggestions future studies to follow-up this work.

7.2. Main findings

Highlights of main findings:

1. There is significant lack of data on the genetic determinants of iron status on African populations.
2. Wide disparity exists in allele frequencies of single nucleotide polymorphisms (SNPs) associated with iron status between Africans and non-African populations
3. The linkage disequilibrium patterns between common *TMPRSS6* SNPs reported in Europeans differ from that of Africans.
4. Previously reported functional effects of *TF* rs3811647 on transferrin and iron binding capacity replicates in Gambians.
5. *TMPRSS6* rs2235321 influence hepcidin concentrations.
6. We observed reverse trend on the effects of *TMPRSS6* variant alleles on hepcidin, contradicting previous reports.

7.3. Discussion of main findings

In chapter 3, a total of 64 studies were retrieved from the literature search, but only five were conducted in Africa. This finding highlights a significant gap in the availability of data on the genetic determinants of iron imbalance in Africa (**Chapter 3, Fig 1**). This demonstrates a critical need to conduct human genetic research in Africa to understand the contribution of genetic risk factors on the high burden of anaemia in African populations. Also, one of the key findings in this study was the wide disparity in allele frequency for most of the SNPs associated with iron status across global populations (**Chapter 3, Supplemental Table 1**). For example, *TMPRSS6* rs855791, the most widely reported SNP that is associated with iron-refractory iron deficiency anaemia and impaired iron status indicators, has an average global minor allele frequency (MAF) of 40%. However, the MAF of *TMPRSS6* rs855791 in Africans and Gambians is 10% and 7% respectively in the 1000 Genomes Project.

Another significant finding from this study was the differences in linkage equilibrium (LD) observed between the common *TMPRSS6* SNPs (rs855791 and rs2235321 or rs4820268) in different populations. *TMPRSS6* rs855791 and rs2235321 were previously reported to be in high LD in Europeans (D'= 1.0, r^2=0.44)[1]. Similarly, rs4820268 was reported to be in high LD with rs855791 in Italians (r^2=0.81) [2]. However, we found that rs855791 is in low LD with rs2235321 and rs4820268 both in the pan-African data in the 1000 Genomes project and in the Gambian population we studied (**Chapter 3, Figure 4**). The differences in LD between SNPs in different populations demonstrates that genetic results from one population may not be transferrable to others [3].

In **Chapter 4**, we sought to investigate the whether the common *TMPRSS6* and *TF* SNPs will have an effect on iron status indicators. We did not detect any effect

of *TMPRSS6* genotype combinations or allele risk scores with any iron biomarker. This result indicated that these *TMPRSS6* SNPs might not have an impact on iron status in Africans, unlike the previously reported associations in non-African populations [4–8]. However, we found *TMPRSS6* rs2235321 to have an effect on plasma hepcidin concentration with a stronger impact on individuals with lower haemoglobin (Hb) and ferritin levels. This suggest that despite being a synonymous variant, *TMPRSS6* rs2235321 may influence hepcidin regulation of iron homeostasis.

Also, we found *TF* rs3811647 to have a strong association with transferrin and transferrin binding (TSAT and UIBC). *TF* rs3811647 homozygotes (AA) had a 20% higher transferrin than the GG carriers. Our results indicate that this SNP is associated with transferrin binding of iron but not iron itself. This finding is in agreement with previous studies in both Europeans [9,10] and Africans [11], on the association between *TF* rs3811647 and iron status. Further research could assess how the *TF* rs3811647 influences the functionality of transferrin and how this impact on the risk of iron deficiency in Africans.

In **Chapter 6**, we conducted a recall-by-genotype (RbG) study to test the hypothesis that individuals carrying the variant alleles at the common *TMPRSS6* SNPs might have less effective response to oral iron supplementation compared to the carriers of the major alleles. However, we failed to detect any effect of variants alleles on the response to oral iron ingestion in this population. Unexpectedly, for each of the three SNPs studies, individuals who are homozygotes for the minor alleles (rs2235321 AA, rs855791 AA and rs4820268 GG) had lower hepcidin concentrations compared to the homozygous major allele carriers (**Chapter 6, Figure 6**). These results are at variance with the previous reports on these SNPs, and they contradict what we expected to find. We expected that the carriage of minor alleles at *TMPRSS6* rs2235321, rs855791

and rs4820268 would lead to elevated hepcidin, due to impaired *TMPRSS6* function caused by these variants [12,13]. *TMPRSS6* rs855791 A allele is associated with elevated hepcidin accompanied by low TSAT and serum in Europeans [7]. *TMPRSS6* rs2235321 A and rs4820628 G alleles are associated with the risk of low iron status including in Africans [4,14–16]. However, we did not find previous reports on the effects of rs2235321 or rs4820268 on hepcidin concentration or response to iron supplementation.

One significant finding from the RbG study is that we replicated the effects of rs2235321 on hepcidin we observed in the cross-sectional study (**Chapter 4**). In both studies, the carriers of the homozygous A allele have lower hepcidin levels than the GG carriers (**Chapter 4 Figure 2** and **Chapter 6 Figure 5A**). These results demonstrate a possible effect of *TMPRSS6* rs2235321 on hepcidin, which requires further investigation, particularly given the high MAF of this SNP in Africans (41%) and 43% in Gambians.

Another significant finding from this study is the identification of a genotype group with individuals that had a static hepcidin concentration despite ingesting a bolus dose of oral iron. The hepcidin level in all the genotype groups rose after the oral iron dose, except the group consisting of individuals with rs2235321 AG and rs855791 AG simultaneously (**Chapter 6, Figure 5**). The hepcidin concentration in this group remained unchanged whilst the plasma iron concentration rose after iron ingestion. These results may suggest that there might be an alternate pathway that is responsible for iron regulation at the enterocytes or that highly bioavailable iron evaded the hepcidin-mediated iron regulation at the enterocytes. Further research with larger sample size involving carriers of this genotype, and in different population groups might provide more insight into these results.

7.4. Limitations implications

The candidate gene approach and the recall-by-genotype study design was made possible by the recent surge in GWAS to search for genetic variants that influence iron status. As mentioned in the methods of the different chapters, we chose the candidate SNPs based on the previous associations between these variants and various iron status indicators. However, as we note in the study presented in **Chapter 3**, the bulk of these studies were conducted in Europeans and a few on Asians, and we found only five studies that were conducted in Africa.

The primary purpose of employing the recall-by-genotype study (RbG) design was to ensure that we focus on individuals with the SNPs of interest rather than using the standard case-control study design. Although we ensured that each genotype group was unique, we could not account for the presence of variants in other genes that influence iron status. For example, *TF* rs3811647 has a strong effect on transferrin and its ability to bind iron, as identified in previous research and our study in **Chapter 5**. Unfortunately, we could not account for the confounding effects of *TF* rs3811647 in our RbG study. To eliminate the possible confounding effects of other genetic variants would require a larger pre-genotyped and traceable population. Larger sample size would require additional resources and time, which is beyond the scope of this PhD.

Another limitation of this study design arises when it is applied to study low-frequency variants. The low minor allele frequency (MAF) of *TMPRSS6* rs855791 in our study population and Africans impacted our study. We could not study homozygotes for the minor allele rs855791 (AA). We only had heterozygotes (AG) and homozygotes for the major allele (GG) to include. Therefore, the lack of individuals with rs855791 AA only was a notable limitation.

7.5. Public health implications and recommendations

In this thesis, I attempted to assess whether common genetic variants previously identified in Europeans and Asians could increase the risk of anaemia and predispose to inadequate response to oral iron supplementation in Africans. This aim was conceived based on previous research findings linking genetic variants within the *TMPRSS6* gene to the risk of iron refractory iron deficiency anaemia (IRIDA) [17,18]. The key features of IRIDA include inappropriately elevated hepcidin for the degree of anaemia and inability to absorb oral iron [18,19]. IRIDA patients usually present with varying degrees of anaemia, ranging from moderate to severe microcytosis with low serum iron and low-normal ferritin [17,20–22].

Given that hepcidin elevation is a classic finding in patients with IRIDA, hepcidin suppression could be identified as a mechanism for enhancing iron absorption. Administration of a hepcidin antagonist could facilitate hepcidin suppression and thereby promote effective iron absorption. Currently, there are several candidate hepcidin antagonists at different stages of clinical development for potential use in the treatment of anaemias characterised with excess hepcidin levels [23–25].

If our hypothesis had been confirmed, carriers of these genetic variants would have potentially been candidates for interventions utilising hepcidin suppression mechanisms. This approach could open the avenue for personalised medicine approaches to anaemia therapy. At the public health level, similar approaches might potentially be useful in settings where the frequency of the genetic variants associated with excess hepcidin is high. Given the high minor allele frequency (MAF) of some of the SNPs studied, we anticipated that identifying sub-population groups carrying these risk alleles might have facilitated population stratification approaches to anaemia control policies. However, we could not confirm that these *TMPRSS6* variants

modulate oral iron absorption or predispose to anaemia in Africans. Thus, at this point, we cannot propose any personalised or public health intervention for the genetic defects in the hepcidin-iron axis.

7.6. Potential future studies

The studies presented in this thesis did not reveal substantial evidence on the effects of common *TMPRSS6* and *TF* SNPs on the risk of anaemia and inadequate oral iron absorption in Africans. However, the results in **Chapter 3** demonstrated that there are several other variants in the hepcidin and iron regulatory genes that are identified as risk factors for low iron status, whose effects on oral iron absorption have not been investigated in Africans. Also, SNPs linked to elevated iron status in Europeans have been reported, but their impact on low iron status have not been thoroughly investigated in Africans either. These include SNPs in the *HFE, SLC40A1* and *HAMP* (the hepcidin gene), *BMP6* and *TFR2* (encoding transferrin receptor 2).

Follow-up research could focus on comprehensively assessing the combine effects risk alleles at all these SNPs on the risk of anaemia. Such studies may be done by expanding our work in **Chapter 5**, where we aggregated the alleles at six *TMPRSS6* SNPs to develop allele risk scores for anaemia. Further studies may enable the development of allele risk scores for anaemia based on all the known SNPs linked to the low iron status from the different iron-related genes. Consequently, the effects of these allele risk scores on various iron status indicators or anaemia and iron deficiency may be validated in different population groups. Such allele scores, also referred to as genetic risk scores, may provide the basis for predicting the occurrence of anaemia and iron deficiency in Africans.

Developing genetic risk scores for complex phenotypes such as BMI [26], blood pressure [27] and type 2 diabetes[28] has been described. Therefore, identifying genetic

risk scores for anaemia and inadequate response to iron supplementation may be useful for improving iron supplementation strategies in different population groups.

Given our finding on the association between *TMPRSS6* rs2235321 and hepcidin, with the effects more pronounced in individuals with low ferritin or haemoglobin (**Chapter 4**), follow-up studies could investigate the effect of this SNP on hepcidin in different population groups with diverse physiological risk factors for impaired iron status. Also, longer-term studies assessing the effect of common genetic variants on iron absorption throughout the routine iron supplementation duration (usually 12 weeks) may provide a better understanding on the role of these variants on the response to oral iron therapy.

Moreover, gene fine-mapping studies may be conducted to determine the exact location of functional variants. Gene fine-mapping refers to the identification of the exact genomic location of functional variants and to precisely determine the causal variant [6181,1829,70]. Similarly, gene expression studies such as expression quantitative trait loci (eQTL) analysis may be conducted to identify new functional variants in the different hepcidin-iron regulatory genes. eQTL refers to the analysis of gene expression in cells or tissues to identify functional variants associated with a phenotype [32]. These types of human genetic research done in Africans may help to provide a clearer insight into the molecular mechanisms underlying iron regulation in African populations.

7.6.1 Specific immediate future research

Based on our findings, I propose follow-up research as follows:

1. To conduct a genome-wide association study (GWAS) on iron status indicators and iron deficiency. This project would use the new H3Africa chip to do a whole-

genome sequencing on the available 8,000 individuals within the Keneba biobank at the MRCG at LSHTM. The H3Africa chip contains novel variants relevant for African populations which were not part of the previous Illumina arrays [33]. Furthermore, the project would use the SNPs identified from the proposed GWAS, and those previously reported from other GWASs to develop genetic risk scores for iron deficiency, using the phenotype data available from the population in the Keneba Biobank. In addition to the SNPs in *TMPRSS6* and *TF*, these genetic risk scores would include variants in other iron regulatory genes such as *HFE*, *DMT1 (SLC11A2)*, *TFR2*, *BMP6*, *SLC40A1*, *HAMP*, and any new loci that may be identified from the Africa-specific discovery GWAS.

2. To use the recall-by-genotype approach to investigate the effects of *TMPRSS6* rs2235321, rs855791, rs4820268 and *TF* rs3811647, and any functional (putative functional) SNP discovered in the GWAS proposed above, on absorption of a more physiological form of iron in anaemic subjects. The studies proposed here may help to elucidate the effects of the common genetic variants in the iron and hepcidin regulatory genes on iron absorption in Africans.

7.7. Conclusions

Finally, the studies presented in this thesis attempted to examine the role of genetic risk factors on impaired iron status response to oral iron supplementation in Africans. We identified an association between *TMPRSS6* 2235321 and hepcidin concentrations, and we replicated the previously reported association between *TF* rs3811647 and transferrin. However, we found a discrepancy between our results and previous observations on the relationship between *TMPRSS6* rs855791 and hepcidin. Therefore, we recommend that more

research is done to examine the reasons for this disparity and to map out whether host genetics risk factors play a role in predisposing Africans to low iron status or hindering the success of iron supplementation strategies. Understanding of the genetic risk factors for anaemia and iron deficiency may facilitate the formulation of population-specific or personalised medicine approaches to anaemia prevention and treatment.

7.8. References

1. Delbini, P. *et al.* Genetic variability of TMPRSS6 and its association with iron deficiency anaemia. *British Journal of Haematology* **151**, 281–284 (2010).
2. Nai, A. *et al.* TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. *Blood* **118**, 4459–4462 (2011).
3. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26–31 (2019).
4. Tanaka, T. *et al.* A genome-wide association analysis of serum iron concentrations. *Blood* **115**, 94–96 (2010).
5. van der Harst, P. *et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369–375 (2012).
6. Seiki, T. *et al.* Association of genetic polymorphisms with erythrocyte traits: Verification of SNPs reported in a previous GWAS in a Japanese population. *Gene* **642**, 172–177 (2018).
7. Nai, A. *et al.* TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. *Blood* **118**, 4459–4462 (2011).
8. Lone, N. M. *et al.* Role of TMPRSS6 rs855791 (T > C) polymorphism in reproductive age women with iron deficiency anemia from Lahore, Pakistan. *Saudi Journal of Biological Sciences* S1319562X20305660 (2020) doi:10.1016/j.sjbs.2020.11.004.
9. Blanco-Rojo, R., Bayele, H. K., Srai, S. K. S. & Vaquero, M. P. Intronic SNP rs3811647 of the human transferrin gene modulates its expression in hepatoma cells. *Nutr Hosp* **27**, 2142–2145 (2012).
10. Manjari, K. S. *et al.* Transferrin (rs3811647) gene polymorphism in iron deficiency anemia. *Mol Cytogenet* **7**, P38 (2014).

11. Gichohi-Wainaina, W. N. *et al.* Common Variants and Haplotypes in the TF, TNF-α, and TMPRSS6 Genes Are Associated with Iron Status in a Female Black South African Population. *The Journal of Nutrition* **145**, 945–953 (2015).
12. Lee, P. Role of Matriptase-2 (TMPRSS6) in Iron Metabolism. *Acta Haematol* **122**, 87–96 (2009).
13. Ramsay, A. J., Hooper, J. D., Folgueras, A. R., Velasco, G. & Lopez-Otin, C. Matriptase-2 (TMPRSS6): a proteolytic regulator of iron homeostasis. *Haematologica* **94**, 840–849 (2009).
14. Delbini, P. *et al.* Genetic variability of TMPRSS6 and its association with iron deficiency anaemia: Short Report. *British Journal of Haematology* **151**, 281–284 (2010).
15. Gichohi-Wainaina, W. N. *et al.* Associations between Common Variants in Iron-Related Genes with Haematological Traits in Populations of African Ancestry. *PLoS ONE* **11**, e0157996 (2016).
16. Chambers, J. C. *et al.* Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat Genet* **41**, 1170–1172 (2009).
17. Sal, E., Keskin, E. Y., Yenicesu, I., Bruno, M. & De Falco, L. Iron-refractory iron deficiency anemia (IRIDA) cases with 2 novel TMPRSS6 mutations. *Pediatric Hematology and Oncology* **33**, 226–232 (2016).
18. Heeney, M. M. & Finberg, K. E. Iron-Refractory Iron Deficiency Anemia (IRIDA). *Hematology/Oncology Clinics of North America* **28**, 637–652 (2014).
19. Finberg, K. E. *et al.* Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA). *Nat Genet* **40**, 569–571 (2008).

20. Pellegrino, R. M. *et al.* Two novel mutations in the tmprss6 gene associated with iron-refractory iron-deficiency anaemia (irida) and partial expression in the heterozygous form. *Br J Haematol* **158**, 668–672 (2012).
21. Bhatia, P., Singh, A., Hegde, A., Jain, R. & Bansal, D. Systematic evaluation of paediatric cohort with iron refractory iron deficiency anaemia (IRIDA) phenotype reveals multiple TMPRSS6 gene variations. *Br J Haematol* **177**, 311–318 (2017).
22. Keskin, E. Y. & Yenicesu, İ. Iron-Refractory Iron Deficiency Anemia. *Tjh* **32**, 1–14 (2015).
23. Renders, L. *et al.* First-in-human Phase I studies of PRS-080#22, a hepcidin antagonist, in healthy volunteers and patients with chronic kidney disease undergoing hemodialysis. *PLoS ONE* **14**, e0212023 (2019).
24. Katsarou, A. & Pantopoulos, K. Hepcidin Therapeutics. *Pharmaceuticals* **11**, 127 (2018).
25. Poli, M., Asperti, M., Ruzzenenti, P., Regoni, M. & Arosio, P. Hepcidin antagonists for potential treatments of disorders with hepcidin excess. *Front. Pharmacol.* **5**, (2014).
26. Fulford, A. J., Ong, K. K., Elks, C. E., Prentice, A. M. & Hennig, B. J. Progressive influence of body mass index-associated genetic markers in rural Gambians. *J Med Genet* **52**, 375–380 (2015).
27. Hoffmann, T. J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* **49**, 54–64 (2017).
28. Udler, M. S., McCarthy, M. I., Florez, J. C. & Mahajan, A. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocrine Reviews* **40**, 1500–1520 (2019).

29. The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

30. Hutchinson, A., Asimit, J. & Wallace, C. Fine-mapping genetic associations. *Human Molecular Genetics* **29**, R81–R88 (2020).

31. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491–504 (2018).

32. Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Phil. Trans. R. Soc. B* **368**, 20120362 (2013).

33. H3ABioNet. An African genotyping array designed by H3ABioNet with the H3Africa Consortium. *H3ABioNet* https://www.h3abionet.org/h3africa-chip (2020).

Chapter 8:

Appendices

8.1. Candidate's contribution to research papers

While I was enrolled in this PhD, I contributed in ongoing research projects conducted by the Nutrition Theme at the MRCG at LSHTM. I have been working with this Research Group prior to starting the PhD and I continue to contribute significantly to research projects outside of my PhD, which resulted in publication of findings in high impact journals. Some of these works involved collaborations with external scientists. Below, I present a listed of the papers I contributed.

1. Prentice S, Jallow AT, Sinjanka E, **Jallow MW**, Sise EA, Kessler NJ, et al. Hepcidin mediates hypoferremia and reduces the growth potential of bacteria in the immediate post-natal period in human neonates. Sci Rep. 2019 Dec;9(1):16596.

2. Armitage AE, Agbla SC, Betts M, Sise EA, **Jallow MW**, Sambou E, et al. Rapid growth is a dominant predictor of hepcidin suppression and declining ferritin in Gambian infants. Haematologica. 2019 Aug;104(8):1542–53.

3. Prentice AM, Bah A, **Jallow MW**, Jallow AT, Sanyang S, Sise EA, et al. Respiratory infections drive hepcidin-mediated blockade of iron absorption leading to iron deficiency anemia in African children. Sci Adv. 2019 Mar;5(3):eaav9020

4. Corbin LJ, Tan VY, Hughes DA, Wade KH, Paul DS, Tansey KE, Butcher F, Dudbridge F, Howson JM, **Jallow MW**, John C, Kingston N, et al. Formalising

recall by genotype as an efficient approach to detailed phenotyping and causal inference. Nat Commun. 2018 Dec;9(1):711. Available at: http://www.nature.com/articles/s41467-018-03109-y.

5. Bah A, Pasricha S-R, **Jallow MW**, Sise EA, Wegmuller R, Armitage AE, et al. Serum Hepcidin Concentrations Decline during Pregnancy and May Identify Iron Deficiency: Analysis of a Longitudinal Pregnancy Cohort in The Gambia. J Nutr. 2017 Jun;147(6):1131–7.

8.2. Ethics approval letters

The Gambia Government/MRC Joint
ETHICS COMMITTEE

C/o MRC Unit: The Gambia, Fajara
P.O. Box 273, Banjul
The Gambia, West Africa
Fax: +220 – 4495919 or 4496513
Tel: +220 – 4495442-6 Ext. 2308
Email: ethics@mrc.gm

8 November 2017

Dr Carla Cerami
Nutrition Theme
MRC Unit The Gambia, Fajara

Dear Dr Cerami

L2017.48, Re Changes to research project: SCC 1429v5. Assessing the effects of risk alleles in hepcidin pathway genes in oral iron absorption.

Thank you for submitting your letter dated 12 September 2017 for consideration by The Gambia Government/MRC Joint Ethics Committee at its meeting held on 27 October 2017.

Our committee is pleased to approve your request.

With best wishes

Yours sincerely

[redacted]

Mr Malamin Sonko
Chairman, Gambia Government/MRC Joint Ethics Committee

Documents submitted for review:

- SCC approval letter – 2 October 2017
- Request letter – 12 September 2017
- Application form for SCC 1429 – 18 September 2017
- Informed Consent Document , version 1.0 – 18 September 2017
- EC approval letter – 28 December 2016

The Gambia Government/MRC Joint
ETHICS COMMITTEE

C/o MRC Unit: The Gambia, Fajara
P.O. Box 273, Banjul
The Gambia, West Africa
Fax: +220 – 4495919 or 4496513
Tel: +220 – 4495442-6 Ext. 2308
Email: ethics@mrc.gm

28 December 2016

Dr Carla Cerami
MRC Unit The Gambia
Keneba

Dear Dr Cerami

L2016.72v1.1, Re SCC 1429v5: Interrogating hepcidin and iron in host-pathogen interaction using a Genes-in-Action study design: Assessing the effects of genetic variations in the hepcidin pathway genes in the response to oral iron supplementation

Thank you for submitting your letter dated 31 October 2016 for consideration by The Gambia Government/MRC Joint Ethics Committee at its meeting held on 16 December 2016.

We are pleased to approve your proposed amendments to the research project.

With best wishes

Yours sincerely

[redacted]
Mr Malamin Sonko
Chairman, Gambia Government/MRC Joint Ethics Committee

Documents submitted for review:-

- SCC approval letter – 9 December 2016
- Updated letter – 6 December 2016
- Request letter – 14 November 2016
- Clinical trial protocol – 13 July 2015
- Informed Consent Documents (pilot and main study), version 2.0 – 16 November 2016
- EC approval letter – 14 June 2016
- SCC approval letter – 21 July 2015
- SCC application form, version 5.0 – 14 November 2016

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Observational / Interventions Research Ethics Committee

Mr Momodou W. Jallow
LSHTM

18 May 2017

Dear Mr Jallow

Submission Title: Assessing the effects of genetic variation in the hepcidin pathway genes in response to oral iron supplementation using a Genes-in-Action study design

LSHTM Ethics Ref: 11679

Thank you for responding to the Observational Committee Chair's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved is as follows:

Document Type	File Name	Date	Version
Local Approval	SCC 1185v2_Hennig_21Apr10-Keneba Biobank GE approval	21/04/2010	2
Information Sheet	KB_BioBank_Informed_Consent_over_18yrs	18/08/2011	4
Information Sheet	Kb-Biobank_information_sheet	21/05/2012	2
Local Approval	L2012.44_Hennig_SCC_Approved_03Oct12	03/10/2012	1
Local Approval	L2012.44_Hennig_SCC_Approved_03Oct12	03/10/2012	1
Local Approval	SCC 1429v4_Hennig_Approved_16Jun16	23/05/2016	4
Information Sheet	GiA-Iron_InfomationSheet&ConsentForm	23/05/2016	1
Information Sheet	GiA_Hypo_InfomationSheet&ConsentForm	23/05/2016	1
Information Sheet	MRCG_Tariff for the Reimbursment of Study Participants_V2-0_25July16[1]	26/07/2016	2
Information Sheet	MRCG_POL-CTS-004_V2-0_Reimbursement of Study Participants[1]	26/07/2016	2
Advertisements	GiA_Studies_flow chart	28/07/2016	1
Investigator CV	CV_Susana_Campino2016	11/11/2016	1
Protocol / Proposal	SCC1429v5_Genes-in-Action-Iron_06-12-2016 APPROVED-GE	06/12/2016	5
Local Approval	SCC1429v5_Genes-in-Action-Iron_06-12-2016 APRROVED-GE	06/12/2016	5
Local Approval	L2016.72v1.1_Cerami_Approved_9Dec16	09/12/2016	1.1
Local Approval	L2016.72v1.1_Cerami_Approved_28Dec16	28/12/2016	1.1
Investigator CV	Carla Cerami CV Jan 2017 MRC	02/01/2017	1
Investigator CV	MWJ_CV_25042017	25/04/2017	2
Information Sheet	SCC1429v5_GiA-Iron_Main-study_InfomationSheet_03052017_Cleaned	03/05/2017	3
Information Sheet	SCC1429v5_GiA-Iron_Main-study_InfomationSheet_27042017_With-trackchanges	03/05/2017	3
Covering Letter	MWJ_LSHTM_Ethics_coverletter_03052017-FINAL	03/05/2017	1
Information Sheet	SCC1429v5_GiA-Iron_Pilot_InfomationSheet_03052017_Cleaned	03/05/2017	3
Information Sheet	SCC1429v5_GiA-Iron_Pilot_InfomationSheet_03052017_with-trackchanges	03/05/2017	3

8.3. Study participants information sheet and consent form

PARTICIPANT INFORMATION SHEET

Genes-in-Action (GiA) Iron Main Study

Version 04 Date 12 September 2017

Study Title: **Assessing the effects of risk alleles in hepcidin pathway genes in modulating the response to iron supplementation**

SCC:	1429v5	Protocol:	N/A

Sponsor & Funder: Medical Research Council Unit ~~The~~ Gambia

1. What is informed consent?

You are invited to take part in a research study. Participating in a research study is not the same as getting regular medical care. The purpose of regular medical care is to improve one's health. The purpose of a research study is to gather information. It is your choice to take part and you can stop any time.

Before you decide you need to understand all information about this study and what it will involve. We will explain to you everything about this study, and you can take time to read the following information or get the information explained to you in your language. Listen carefully and feel free to ask if there is anything that you do not understand. Please ask us any questions that you do not understand. If you would like more information, we are happy to explain this to you more than once.

Please feel free to discuss the study with your spouse, family members or friends, study staff or your doctor or nurse.

If you decide to join the study, you will need to sign or thumbprint a consent form saying you agree to be in the study.

2. Why is this study being done?

As you know many people have "anaemia" (sickness that make someone have less blood), and the number people with this sickness is very high in Gambia. We know that giving people a medicine call "iron" treats and prevents someone from "anaemia", but there is information from people living in Europe and America some information from telling us that not everyone can be treated with iron if they have "anaemia". This is the first time we are doing this study in Gambia. We hope to discover why some people get better when given "iron" to treat "anaemia" but others fail. We hope that one day we could make a medicine that may help us to treat "anaemia" for everyone.

Please ask us any questions that you do not understand. If you would like more information, we are happy to explain this to you more than once.

Please feel free to talk about the study with your doctor or nurse, the study staff, or your family and friends.

3. What does this study involve?

a. You have been contacted to participate based on the information we have about you from previous studies. We get this information from studying something called "genes", which are present in all of us and are responsible for why families look like each other, but different from others. For example, some families are taller or shorter than others. This kind of information is passed from both the father and mother to the children and even onto their grand children. Some of the "genes" may prevent us from getting "anaemia" in the first place, while some other genes may be the reason we get sick when others don't.

b. A fieldworker from MRC Gambia will ask you whether you have any sickness or are currently participating in another MRC study or whether you are pregnant or breastfeeding (for women).

c. This study will take one day to complete, and we will ask you to come to MRC, and you may stay there up to six hours.

d. We will ask to take blood samples at 3 different times during the whole study, and we will give you "iron" tablets to swallow with water.

e. On the study day, we will transport you to the MRC in the morning (around 7am).
   - Once at the MRC we will ask to take 3ml (half teaspoon) of blood.
   - We will then give you two "iron" tablets (400mg ferrous sulfate; 2, 200mg ferrous sulfate) to swallow with water.
   - We will ask to take 3ml (half teaspoon) of blood at 2hours and 5hours after given you the iron tablets.
   - We will measure your height and weight, and your body temperature

If we discover that you are sick and decide that you cannot continue to participate in the study because of that, you will be treated by one of the MRC doctors. If we cannot treat you, we will refer you to the appropriate health facility.

If the research study needs to be stopped, you will be informed and you will have your normal medical care.

4. What will happen to the samples taken in this study?

In order to do the research, we must collect and store blood and health information from people like you. We will do some of the tests right away, but other tests may be done in the future. Once we have done the research we planned, we would like to store your blood and information with other samples that other people have donated.

The blood sample taken from you will be tested by the researchers at MRC Gambia, but they may be sent outside the Gambia to other scientists to do more tests if necessary.

If we find anything that is important for your health, we will contact you and put you in touch with doctors who can help you.

.

5. What harm or discomfort can you expect in the study?

We want to tell you that there are some risks with this study, but this is very little. Most of the time when we take blood, it is safe, but sometimes, when we take blood people feel a bit dizzy or get an infection. There is a chance that you may get a bruise from where we took the blood. If this happen, please let us know and you will be treated.

One potential risk is that "iron supplements" may cause some abdominal discomfort, which may make you feel like wanting to vomit or make you vomit, give you diarrhoea and/or constipation. If you feel any of these conditions, please let us know immediately, you will be treated.

Another potential risk of participating in this study is that information about you may be known to other people who should not have this information. There is a small risk that someone who should not have your information could learn something about you, but this is very unlikely.

6. How the privacy of participants will be protected?

Your blood samples will be stored in freezers in our locked laboratory and your personal information will be on a secure computer. Your name will be replaced by a code (study number) on all your blood samples and information you given us will be removed before we share it with other researchers or transport samples outside The Gambia.

7. Potential benefits

This study will not help you or your family to get better, but we hope that it will benefit others in the future. What we are trying to do is very difficult, and it could take a long time. Whether you decide to join in this study or not it will not affect your treatment in our clinic. You are free to decide whether to join or not at your own will.

8. Will you be compensated for participating in the study?

We will not pay you for participating, but we will pay you back (150 Dalasi) for the time you spend participating in the study during the first day when we ask you to stay at MRC for almost half of the day. We will provide a vehicle to bring you to MRC and back to your home. If you come on your own, we will pay you back the money spent on fares (variable amount).

During your stay at MRC on the first day, we will give you breakfast and lunch, and water to drink.

9. What happens if you refuse to participate in the study or change your mind later?

You are free to participate or not in the study and you have the right to stop participating at anytime without giving a reason. This will not affect the medical care or transportation that you would normally receive.

In case you decide to withdraw your participation before the end of the study, any information already generated from the samples until the time of withdrawal will be used.

10. Who should you contact if you have questions?

If you have any queries or concern you can contact or Momodou Wuri Jallow on **7710693**.

Please feel free to ask any question you might have about the research study.

11. Return of results

When the study is finished, we will share the study's general research findings with you and with your community. We think this may take about 1year or more.
In some situations, the results might be important to your health care. If that occurs, we will contact you to see if you want to learn more.

12. How will personal records remain confidential and who will have access to it?

All information that is collected about you in the course of the study will be kept strictly confidential. Your personal information will only be available to the study team members, your healthcare provider and might be seen by some rightful persons from the Ethics Committee, Government authorities and sponsor.

13. Who has reviewed this study?

This study has been reviewed and approved by a panel of scientists at the Medical Research Council and the Gambia Government/MRC Joint Ethics Committee, which consists of scientists and lay persons to protect your rights and wellbeing.

CONSENT FORM

Participant Identification Number: |__|__|__|__|__|__|__|__|__|__|__|__|__|__|

_____
(Printed name of participant)

☐ I have read the written information **OR**

☐ I have had the information explained to me by study personnel in a language that I understand,

and I

- confirm that my choice to participate is entirely voluntarily,
- confirm that I have had the opportunity to ask questions about this study and I am satisfied with the answers and explanations that have been provided,
- understand that I grant access to data about me to authorised persons described in the information sheet,
- understand that the information about me and sample collected from me will be used to support other research in the future, and may be shared anonymously with other researchers, for their ethically-approved projects,
- have received sufficient time to consider to take part in this study,
- agree to take part in this study.

Tick as appropriate

I agree for my samples to be shipped outside of The Gambia Yes ☐ No ☐

I agree to further research on my samples as described in the information sheet Yes ☐ No ☐

Participant's signature/ thumbprint* ☐

_____
Date (dd/mmm/yyyy) Time (24hr)

Printed name of witness* _____

Printed name of person obtaining consent _____

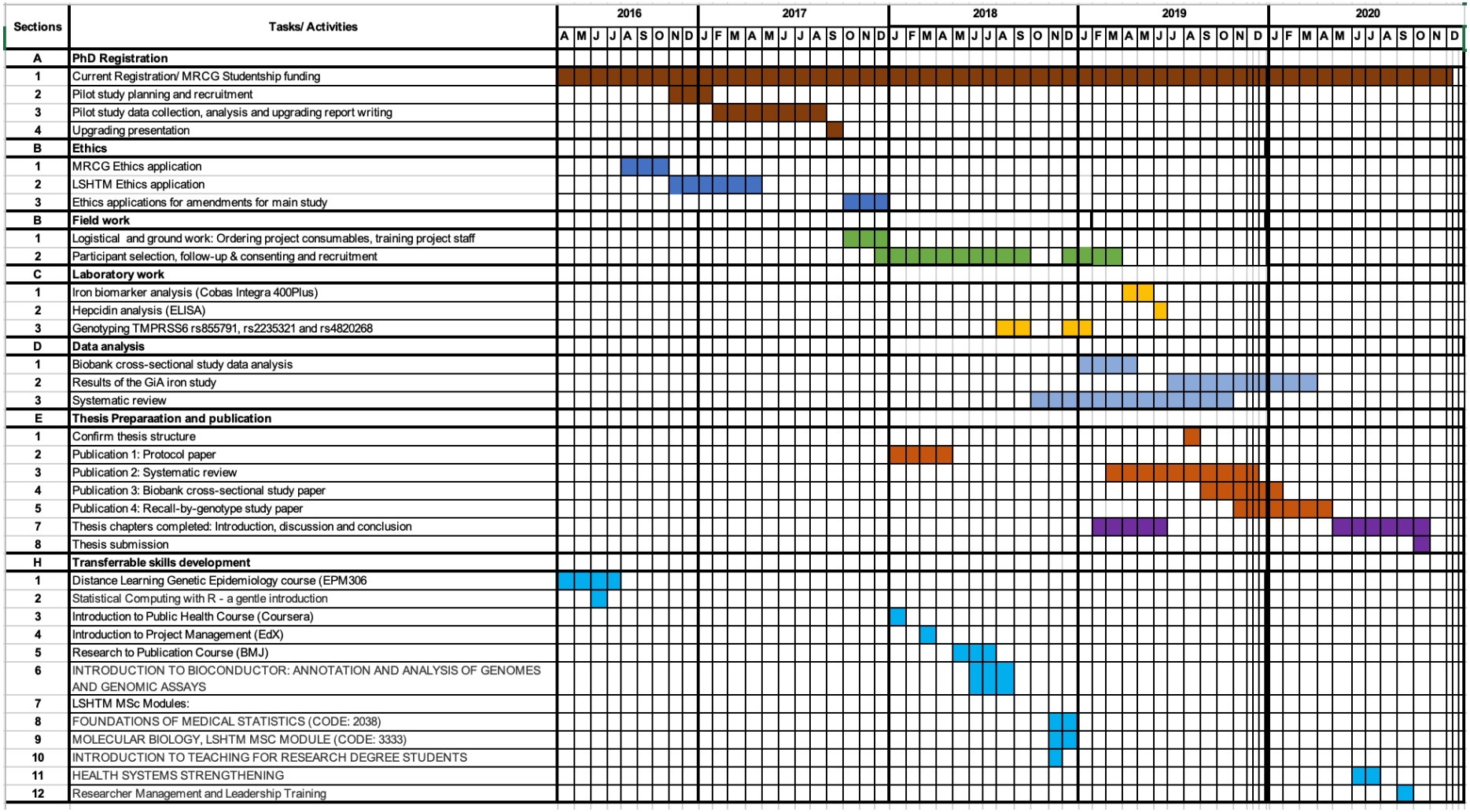
I attest that I have explained the study information accurately in _____ to, and was understood to the best of my knowledge by, the participant. He/she has freely given consent to participate *in the presence of the above named witness (where applicable).

Signature of person obtaining consent

_____
Date (dd/mmm/yyyy) Time (24hr)

** Only required if the participant is unable to read or write.*

8.4. PhD Timeline



Sections	Tasks/ Activities
A	**PhD Registration**
1	Current Registration/ MRCG Studentship funding
2	Pilot study planning and recruitment
3	Pilot study data collection, analysis and upgrading report writing
4	Upgrading presentation
B	**Ethics**
1	MRCG Ethics application
2	LSHTM Ethics application
3	Ethics applications for amendments for main study
B	**Field work**
1	Logistical and ground work: Ordering project consumables, training project staff
2	Participant selection, follow-up & consenting and recruitment
C	**Laboratory work**
1	Iron biomarker analysis (Cobas Integra 400Plus)
2	Hepcidin analysis (ELISA)
3	Genotyping TMPRSS6 rs855791, rs2235321 and rs4820268
D	**Data analysis**
1	Biobank cross-sectional study data analysis
2	Results of the GiA iron study
3	Systematic review
E	**Thesis Preparaation and publication**
1	Confirm thesis structure
2	Publication 1: Protocol paper
3	Publication 2: Systematic review
4	Publication 3: Biobank cross-sectional study paper
5	Publication 4: Recall-by-genotype study paper
7	Thesis chapters completed: Introduction, discussion and conclusion
8	Thesis submission
H	**Transferrable skills development**
1	Distance Learning Genetic Epidemiology course (EPM306
2	Statistical Computing with R - a gentle introduction
3	Introduction to Public Health Course (Coursera)
4	Introduction to Project Management (EdX)
5	Research to Publication Course (BMJ)
6	INTRODUCTION TO BIOCONDUCTOR: ANNOTATION AND ANALYSIS OF GENOMES AND GENOMIC ASSAYS
7	LSHTM MSc Modules:
8	FOUNDATIONS OF MEDICAL STATISTICS (CODE: 2038)
9	MOLECULAR BIOLOGY, LSHTM MSC MODULE (CODE: 3333)
10	INTRODUCTION TO TEACHING FOR RESEARCH DEGREE STUDENTS
11	HEALTH SYSTEMS STRENGTHENING
12	Researcher Management and Leadership Training

8.5. Pictures taken during the study with permission to reproduce obtained from the participants and staff.

Study participants

The study team